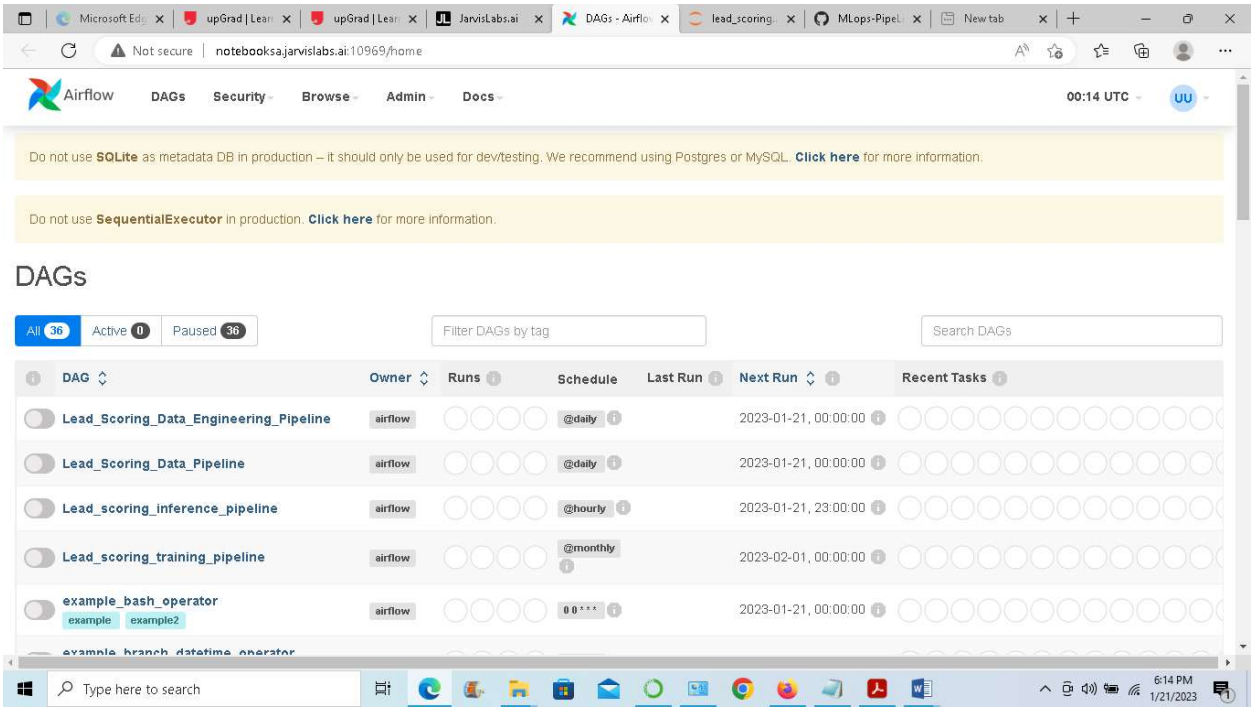
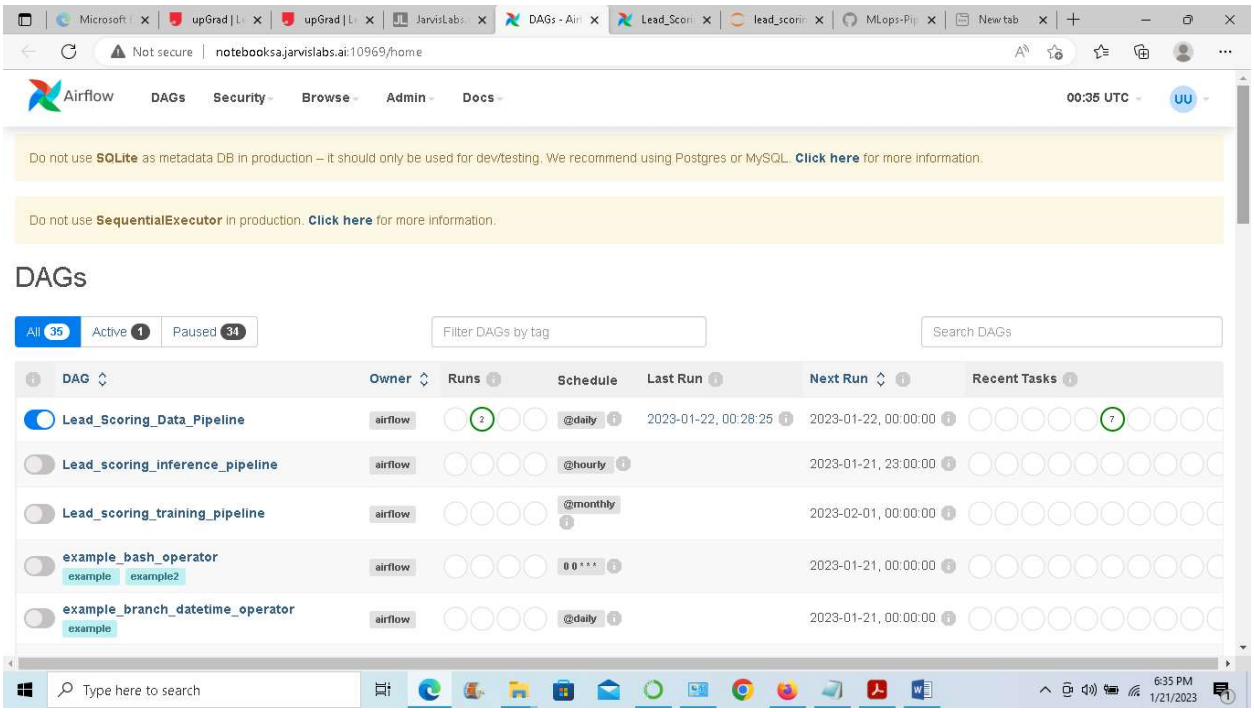
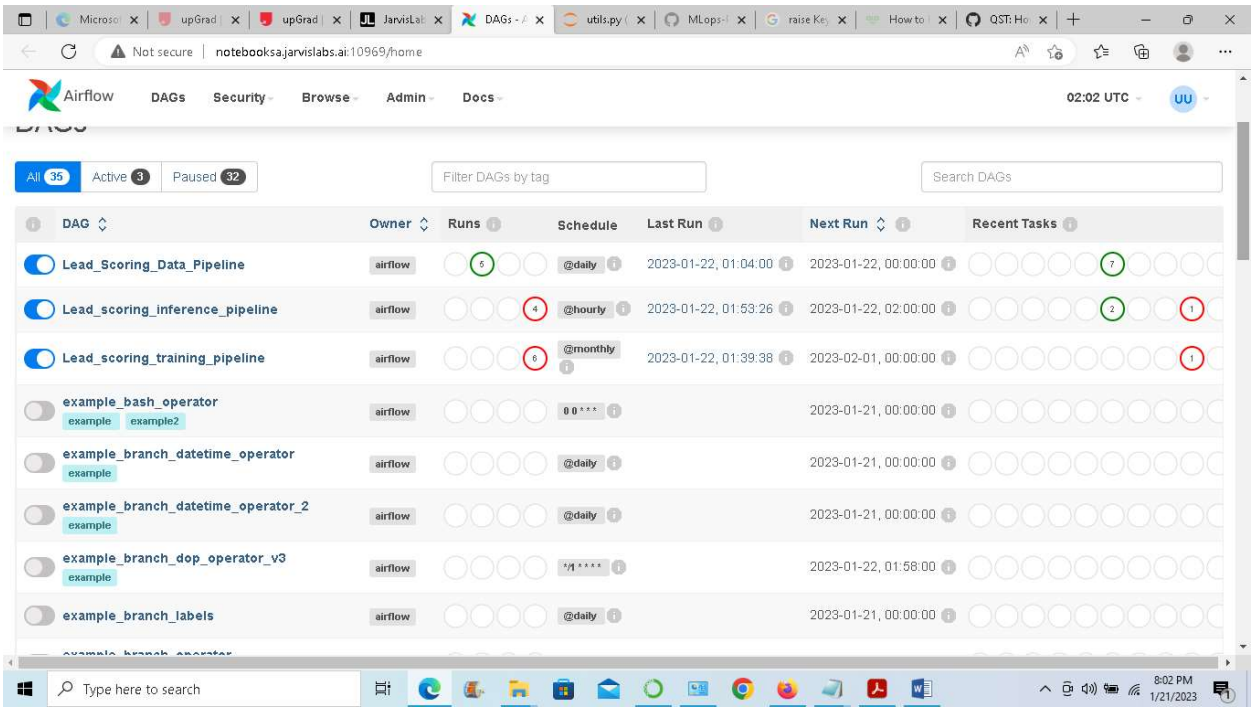
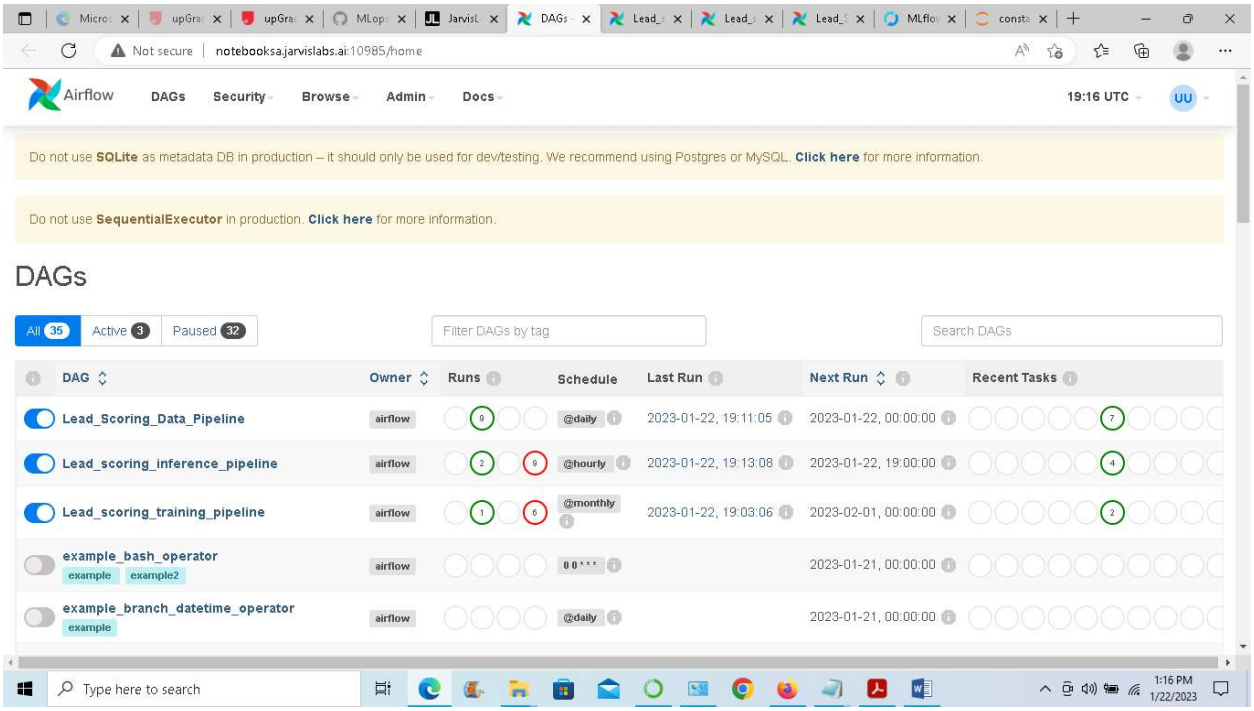
Kavi Prakash Manoharan MLOps Project Submission Using Airflow

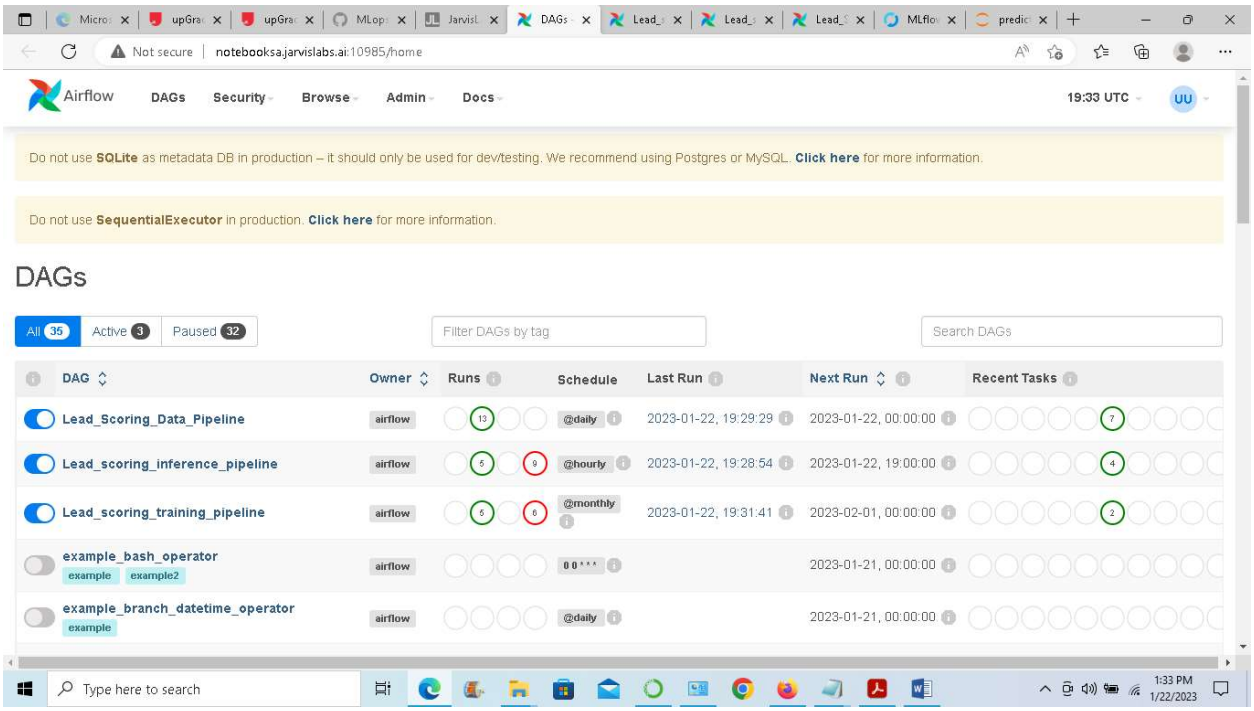
Airflow UI:



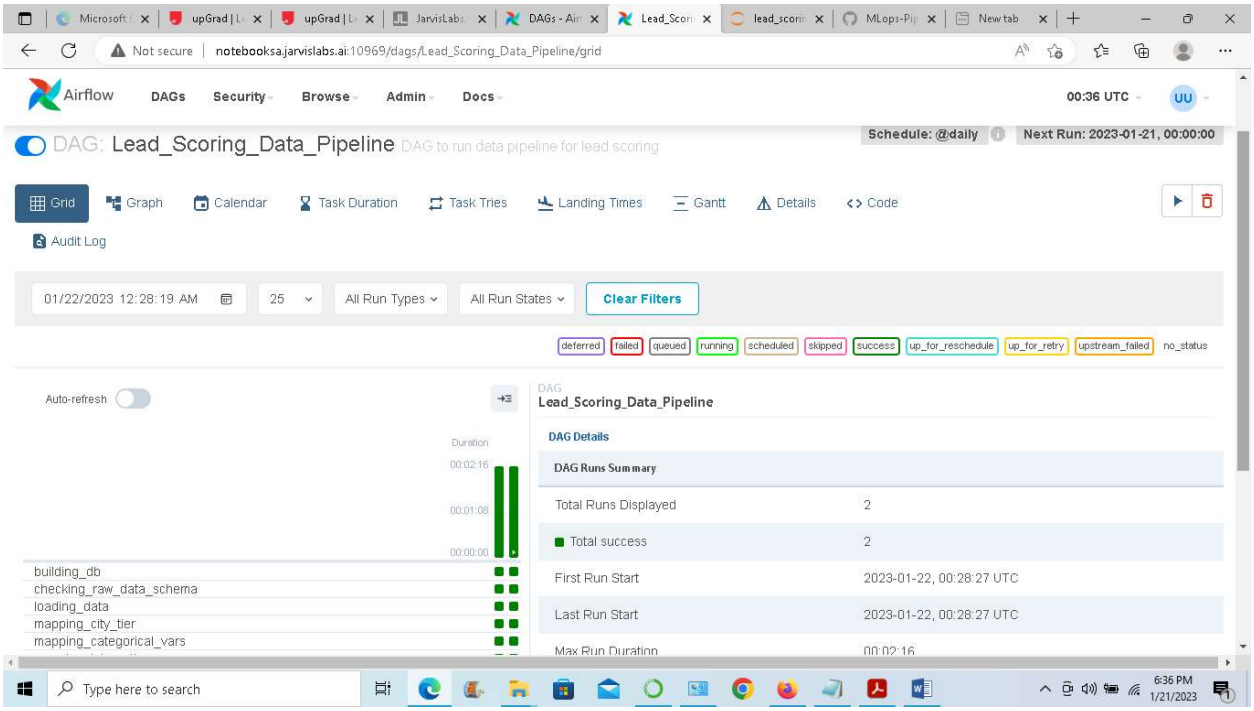


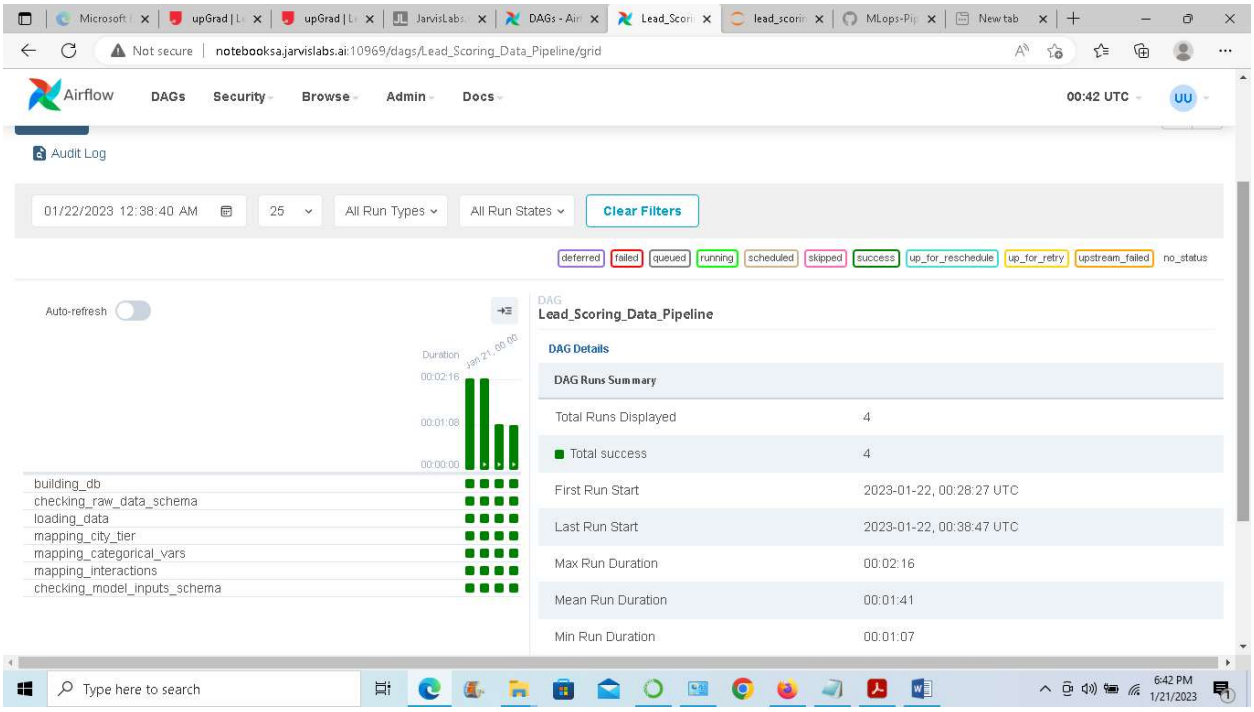

Not secure
notebooksa.jarvislabs.ai:10969/home
Airflow
DAGs
Security
Browse
Admin
Docs
02:02 UTC
All 35
Active 3
Paused 32
Filter DAGs by tag
Search DAGs
DAG
Owner
Runs
Schedule
Last Run
Next Run
Recent Tasks
Lead_Scoring_Data_Pipeline
airflow
@daily
2023-01-22, 01:04:00
2023-01-22, 00:00:00
Lead_scoring_inference_pipeline
airflow
@hourly
2023-01-22, 01:53:26
2023-01-22, 02:00:00
Lead_scoring_training_pipeline
airflow
@monthly
2023-01-22, 01:39:38
2023-02-01, 00:00:00
example_bash_operator
example
example2
airflow
0 0 * * *
2023-01-21, 00:00:00
example_branch_datetime_operator
example
airflow
@daily
2023-01-21, 00:00:00
example_branch_datetime_operator_2
example
airflow
@daily
2023-01-21, 00:00:00
example_branch_dop_operator_v3
example
airflow
*/1 * * * *
2023-01-22, 01:58:00
example_branch_labels
airflow
@daily
2023-01-21, 00:00:00
Type here to search
8:02 PM
1/21/2023

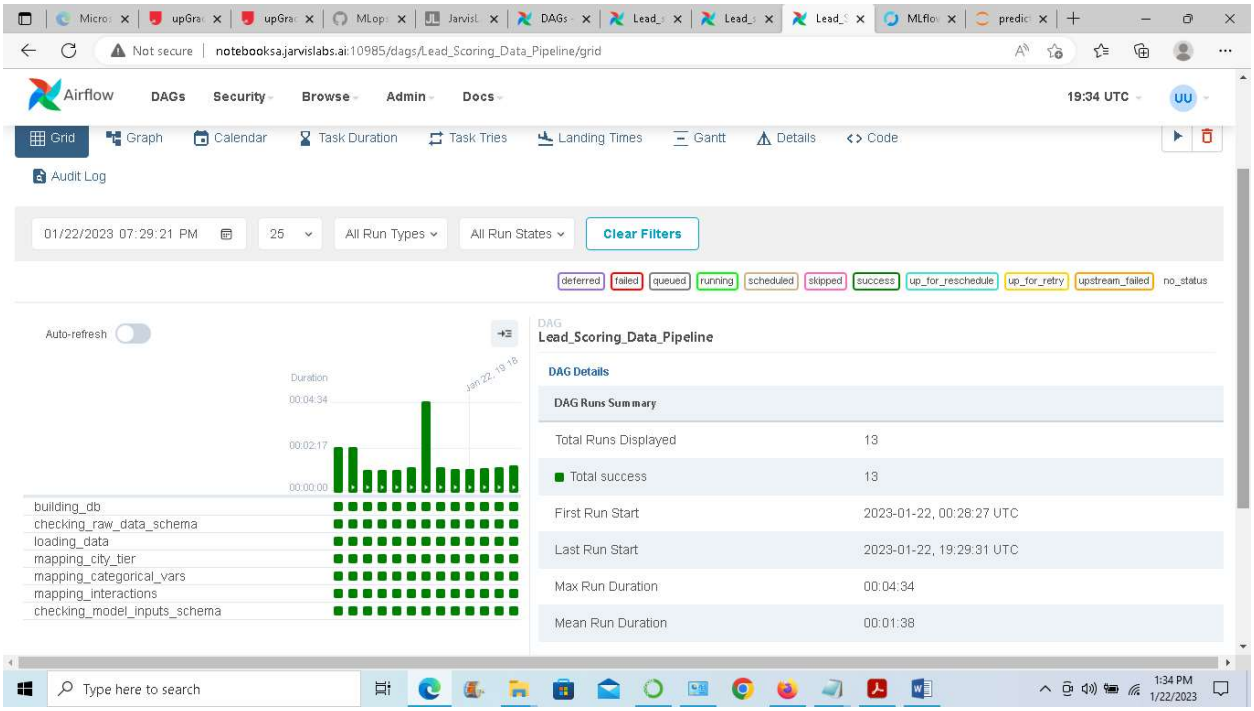

Not secure
notebooksa.jarvislabs.ai:10985/home
Airflow
DAGs
Security
Browse
Admin
Docs
19:16 UTC
Do not use SQLite as metadata DB in production – it should only be used for dev/testing. We recommend using Postgres or MySQL. Click here for more information.
Do not use SequentialExecutor in production. Click here for more information.
DAGs
All 35
Active 3
Paused 32
Filter DAGs by tag
Search DAGs
DAG
Owner
Runs
Schedule
Last Run
Next Run
Recent Tasks
Lead_Scoring_Data_Pipeline
airflow
@daily
2023-01-22, 19:11:05
2023-01-22, 00:00:00
Lead_scoring_inference_pipeline
airflow
@hourly
2023-01-22, 19:13:08
2023-01-22, 19:00:00
Lead_scoring_training_pipeline
airflow
@monthly
2023-01-22, 19:03:06
2023-02-01, 00:00:00
example_bash_operator
example
example2
airflow
0 0 * * *
2023-01-21, 00:00:00
example_branch_datetime_operator
example
airflow
@daily
2023-01-21, 00:00:00
Type here to search
1:16 PM
1/22/2023

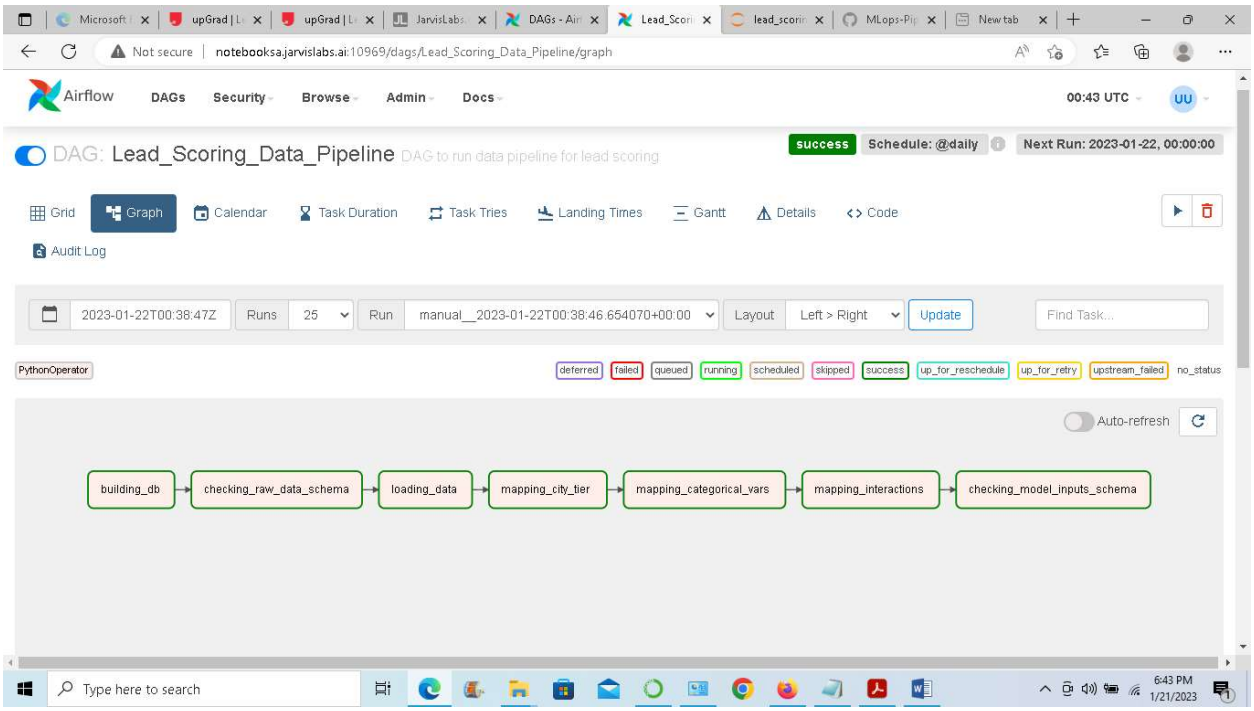


Data PipeLine Run:

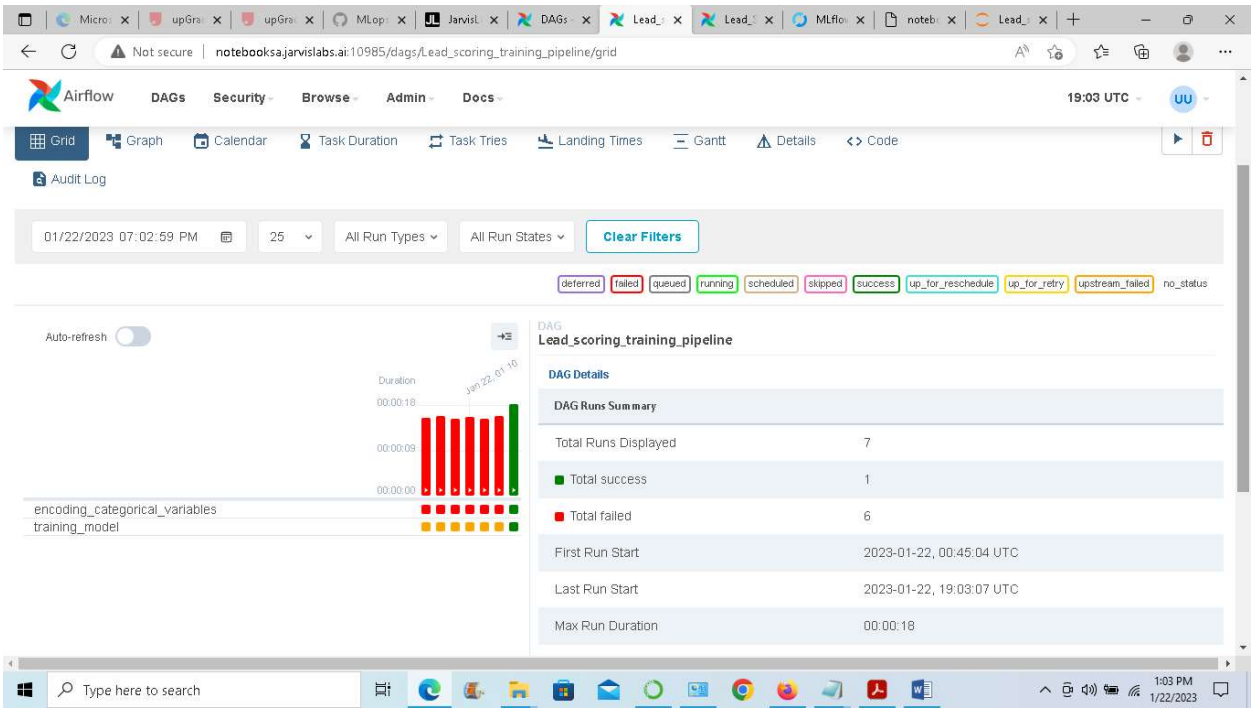


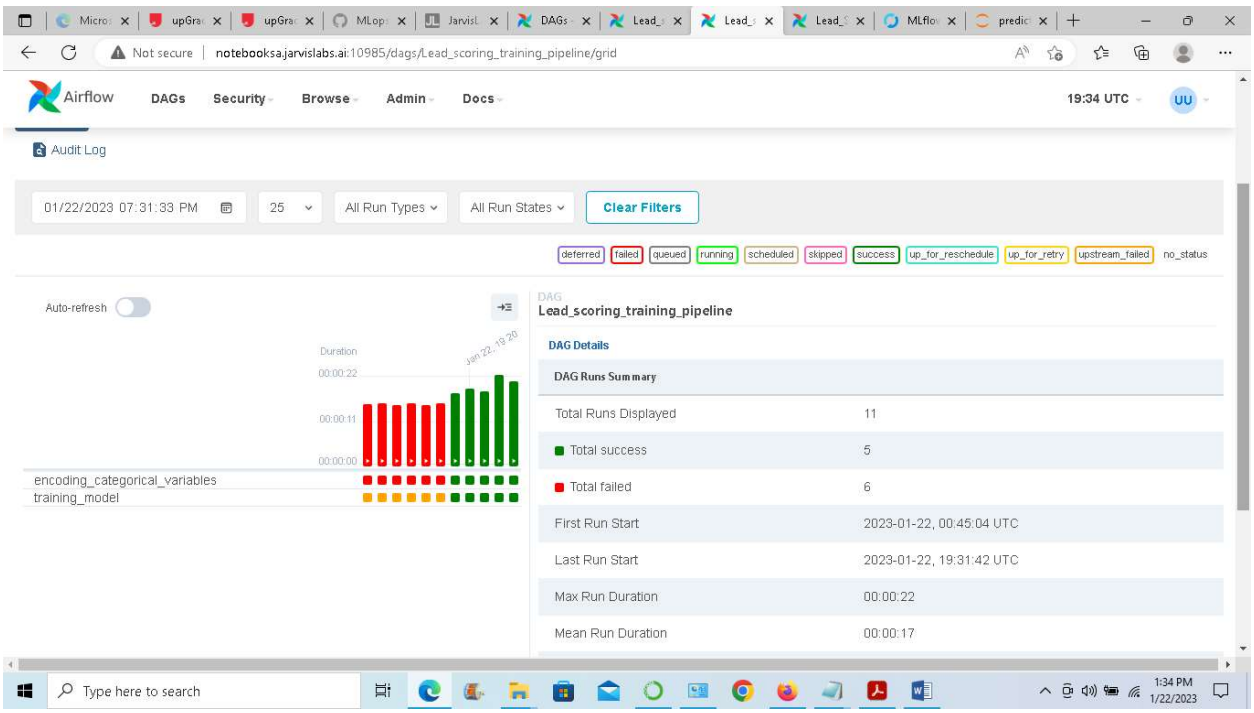

Not secure
notebooksa.jarvislabs.ai:10969/dags/Lead_Scoring_Data_Pipeline/grid
Airflow
DAGs
Security
Browse
Admin
Docs
00:42 UTC
Audit Log
01/22/2023 12:38:40 AM
25
All Run Types
All Run States
Clear Filters
deferred
failed
queued
running
scheduled
skipped
success
up_for_reschedule
up_for_retry
upstream_failed
no_status
Auto-refresh
DAG
Lead_Scoring_Data_Pipeline
Duration
00:02:16
00:01:08
00:00:00
building_db
checking_raw_data_schema
loading_data
mapping_city_tier
mapping_categorical_vars
mapping_interactions
checking_model_inputs_schema
DAG Details
DAG Runs Summary
Total Runs Displayed 4
Total success 4
First Run Start 2023-01-22, 00:28:27 UTC
Last Run Start 2023-01-22, 00:38:47 UTC
Max Run Duration 00:02:16
Mean Run Duration 00:01:41
Min Run Duration 00:01:07
Type here to search
6:42 PM
1/21/2023

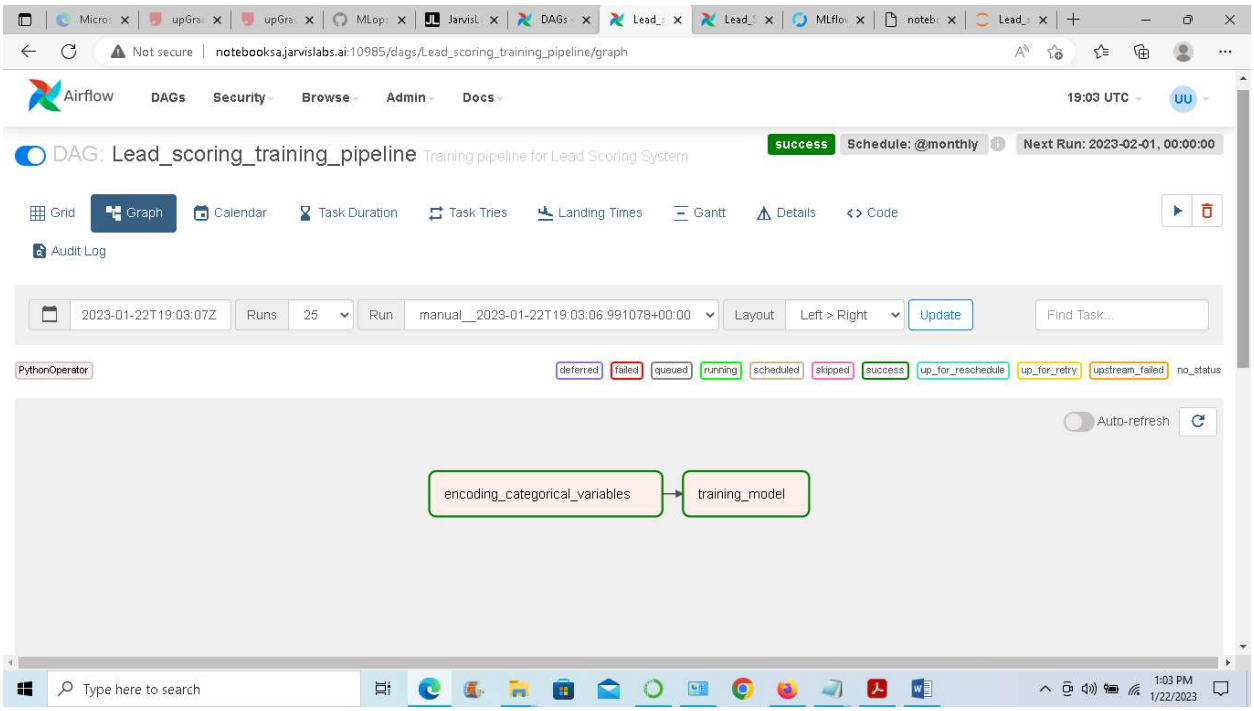

Not secure
notebooksa.jarvislabs.ai:10985/dags/Lead_Scoring_Data_Pipeline/grid
Airflow
DAGs
Security
Browse
Admin
Docs
19:34 UTC
Grid
Graph
Calendar
Task Duration
Task Tries
Landing Times
Gantt
Details
Code
Audit Log
01/22/2023 07:29:21 PM
25
All Run Types
All Run States
Clear Filters
deferred
failed
queued
running
scheduled
skipped
success
up_for_reschedule
up_for_retry
upstream_failed
no_status
Auto-refresh
DAG
Lead_Scoring_Data_Pipeline
Duration
00:04:34
00:02:17
00:00:00
building_db
checking_raw_data_schema
loading_data
mapping_city_tier
mapping_categorical_vars
mapping_interactions
checking_model_inputs_schema
DAG Details
DAG Runs Summary
Total Runs Displayed 13
Total success 13
First Run Start 2023-01-22, 00:28:27 UTC
Last Run Start 2023-01-22, 19:29:31 UTC
Max Run Duration 00:04:34
Mean Run Duration 00:01:38
Type here to search
1:34 PM
1/22/2023


Not secure
notebooksa.jarvislabs.ai:10969/dags/Lead_Scoring_Data_Pipeline/graph
Airflow
DAGs
Security
Browse
Admin
Docs
00:43 UTC
DAG: Lead_Scoring_Data_Pipeline
DAG to run data pipeline for lead scoring
success
Schedule: @daily
Next Run: 2023-01-22, 00:00:00
Grid
Graph
Calendar
Task Duration
Task Tries
Landing Times
Gantt
Details
Code
Audit Log
2023-01-22T00:38:47Z
Runs
25
Run
manual__2023-01-22T00:38:46.654070+00:00
Layout
Left > Right
Update
Find Task...
PythonOperator
deferred
failed
queued
running
scheduled
skipped
success
up_for_reschedule
up_for_retry
upstream_failed
no_status
Auto-refresh
building_db
checking_raw_data_schema
loading_data
mapping_city_tier
mapping_categorical_vars
mapping_interactions
checking_model_inputs_schema
Type here to search
6:43 PM
1/21/2023

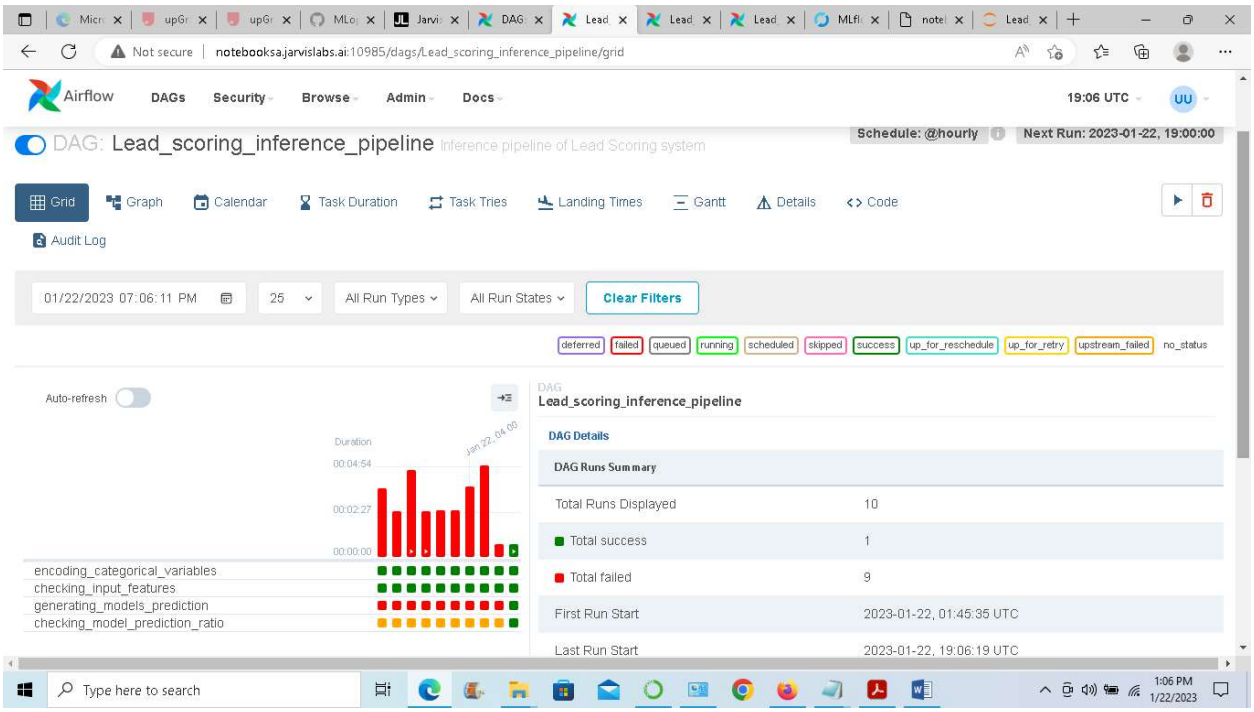
Training Pipeline

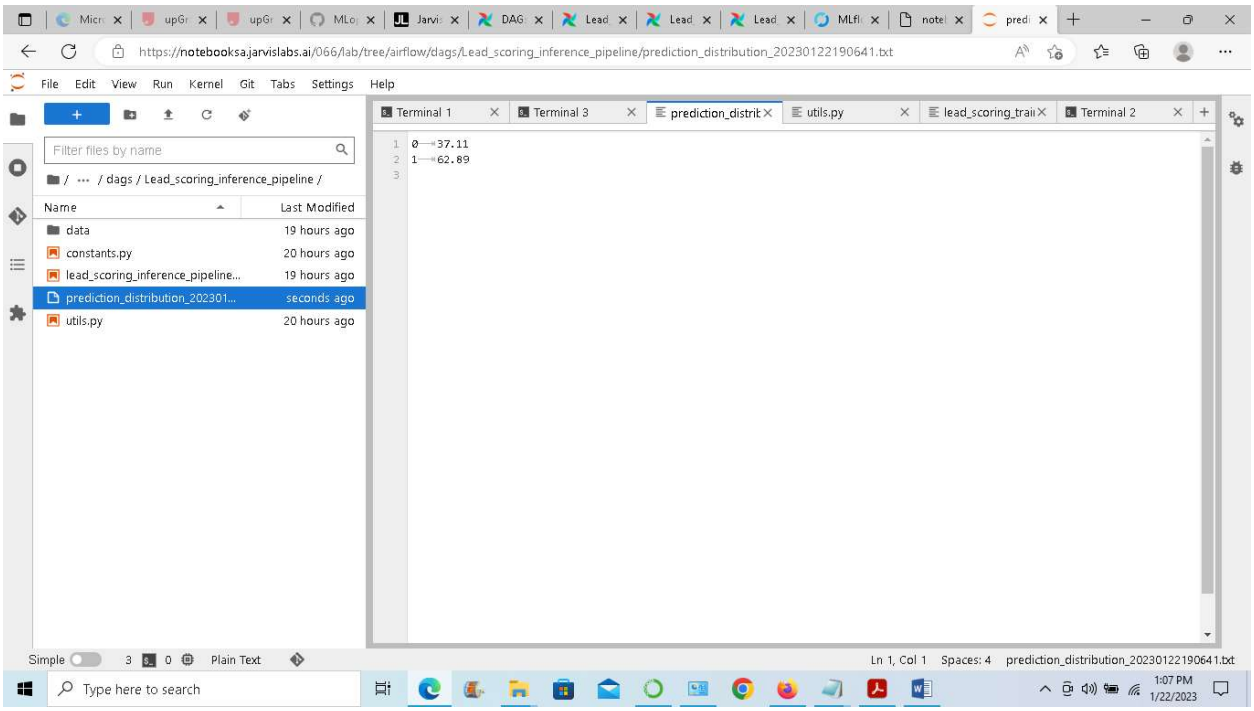


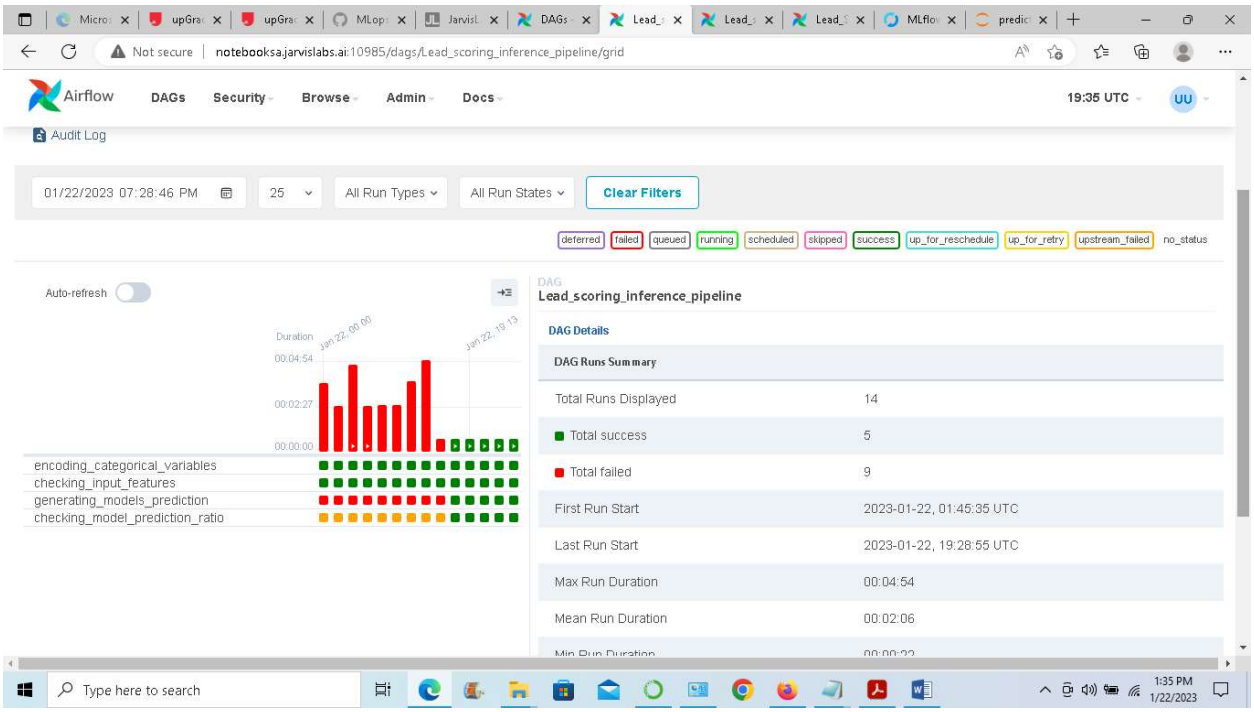


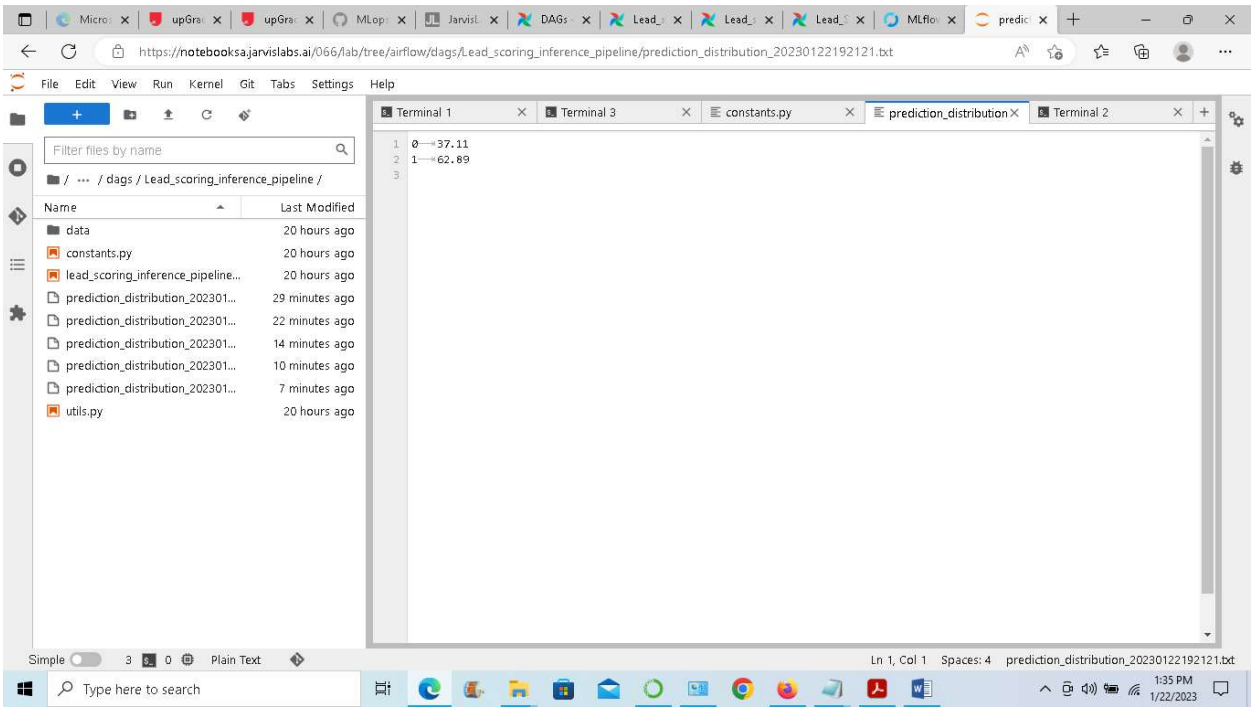


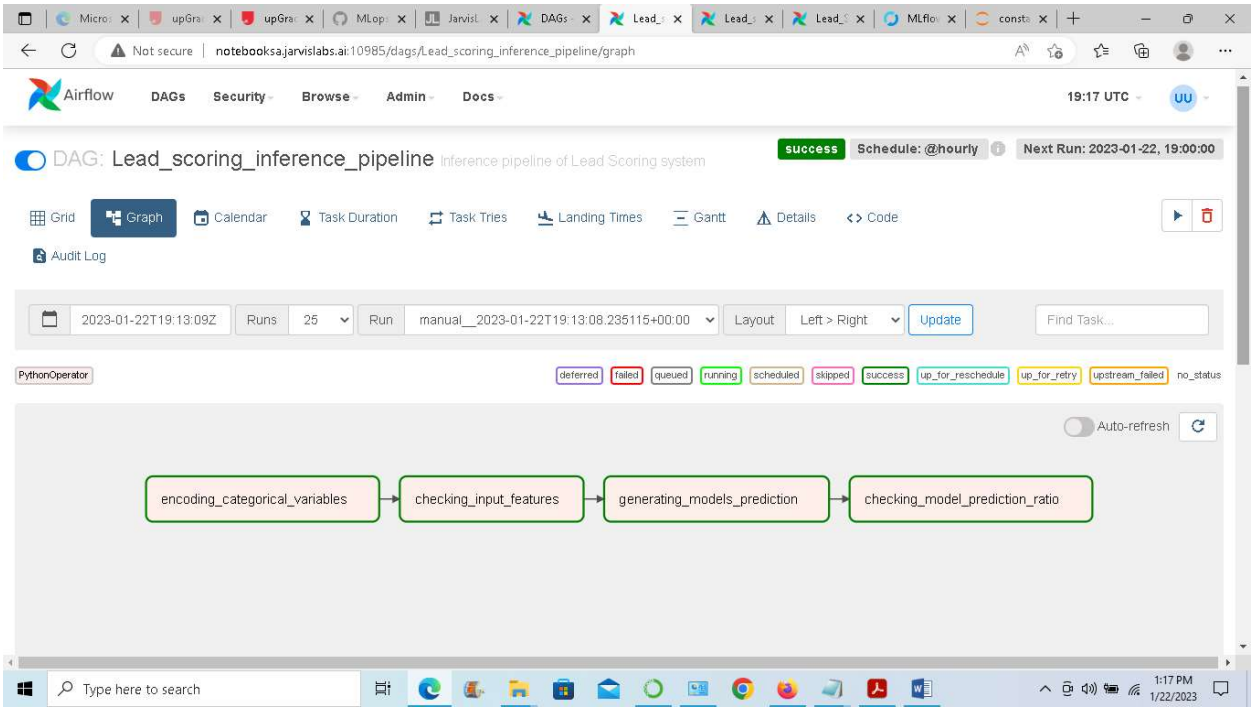
Inference Pipeline



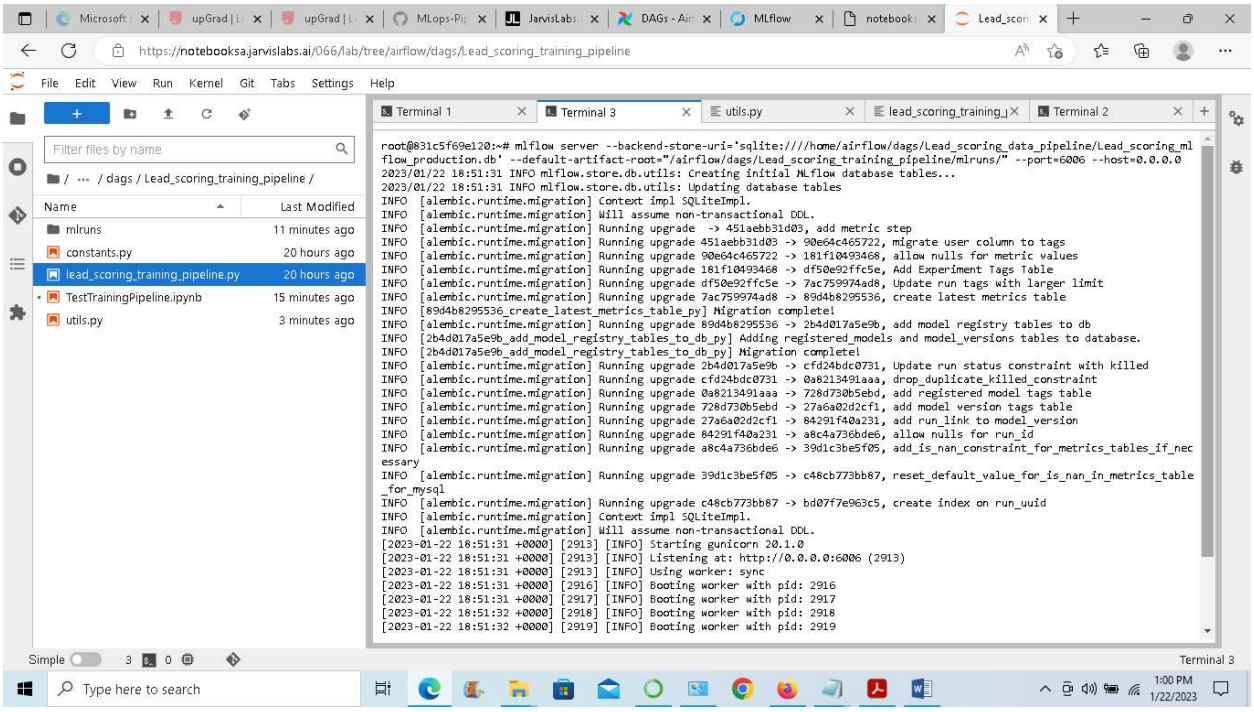

https://notebooksa.jarvislabs.ai/066/lab/tree/airflow/dags/Lead_scoring_inference_pipeline/prediction_distribution_20230122190641.txt
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ dags / Lead_scoring_inference_pipeline /
Name
Last Modified
data 19 hours ago
constants.py 20 hours ago
lead_scoring_inference_pipeline... 19 hours ago
prediction_distribution_202301... seconds ago
utils.py 20 hours ago
Terminal 1
Terminal 3
prediction_distrit
utils.py
lead_scoring_trail
Terminal 2
0 37.11
1 62.89
Simple
Plain Text
Ln 1, Col 1
Spaces: 4
prediction_distribution_20230122190641.txt
Type here to search
1:07 PM
1/22/2023

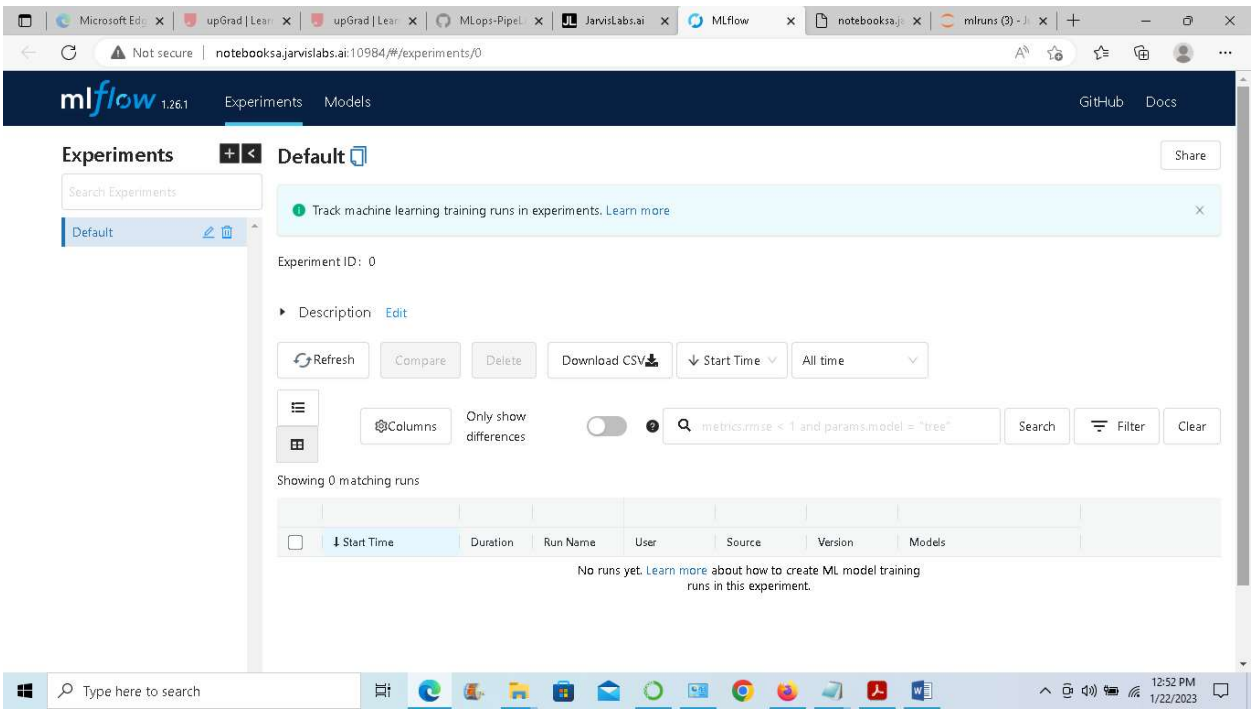

Not secure
notebooksa.jarvislabs.ai:10985/dags/Lead_scoring_inference_pipeline/grid
Airflow
DAGs
Security
Browse
Admin
Docs
19:35 UTC
UU
Audit Log
01/22/2023 07:28:46 PM
25
All Run Types
All Run States
Clear Filters
deferred
failed
queued
running
scheduled
skipped
success
up_for_reschedule
up_for_retry
upstream_failed
no_status
Auto-refresh
Duration
00:04:54
00:02:27
00:00:00
encoding_categorical_variables
checking_input_features
generating_models_prediction
checking_model_prediction_ratio
DAG
Lead_scoring_inference_pipeline
DAG Details
DAG Runs Summary
Total Runs Displayed 14
Total success 5
Total failed 9
First Run Start 2023-01-22, 01:45:35 UTC
Last Run Start 2023-01-22, 19:28:55 UTC
Max Run Duration 00:04:54
Mean Run Duration 00:02:06
Type here to search
1:35 PM
1/22/2023

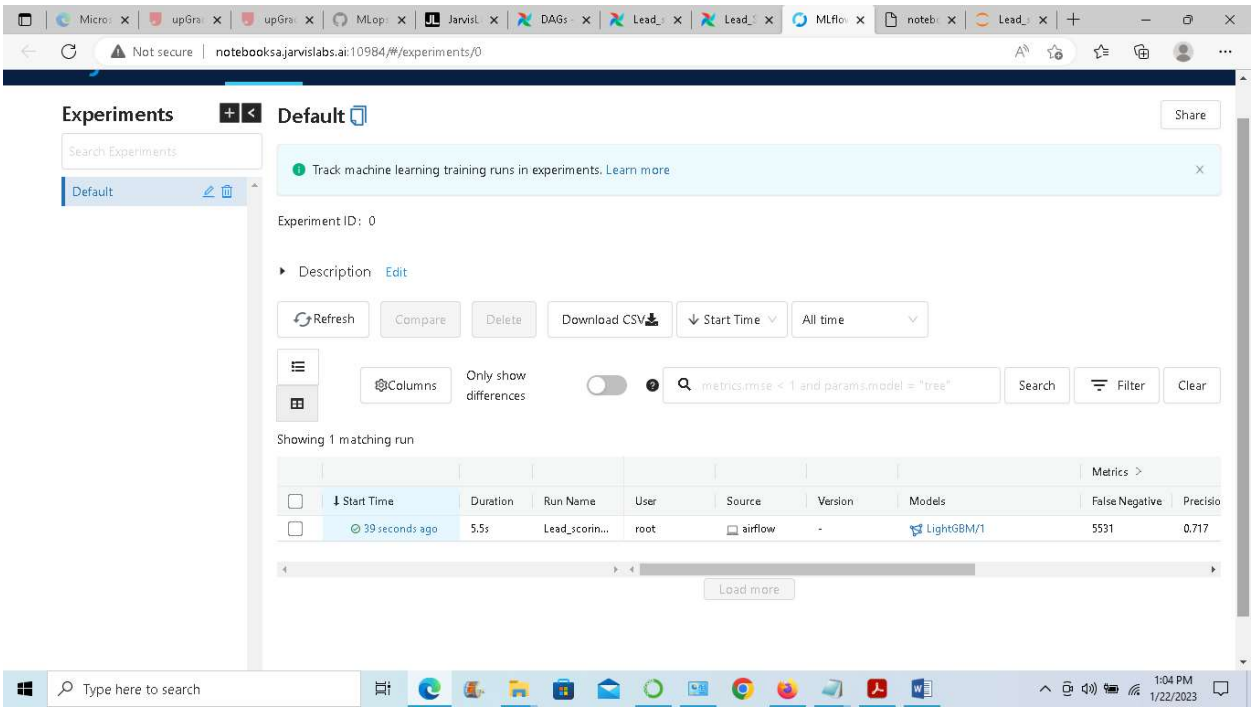

https://notebooksa.jarvislabs.ai/066/lab/tree/airflow/dags/Lead_scoring_inference_pipeline/prediction_distribution_20230122192121.txt
File Edit View Run Kernel Git Tabs Settings Help
Terminal 1
Terminal 3
constants.py
prediction_distribution
Terminal 2
Filter files by name
/ dags / Lead_scoring_inference_pipeline /
Name
Last Modified
data 20 hours ago
constants.py 20 hours ago
lead_scoring_inference_pipeline... 20 hours ago
prediction_distribution_202301... 29 minutes ago
prediction_distribution_202301... 22 minutes ago
prediction_distribution_202301... 14 minutes ago
prediction_distribution_202301... 10 minutes ago
prediction_distribution_202301... 7 minutes ago
utils.py 20 hours ago
0→37.11
1→62.89
Simple
Plain Text
Ln 1, Col 1
Spaces: 4
prediction_distribution_20230122192121.txt
Type here to search
1:35 PM
1/22/2023

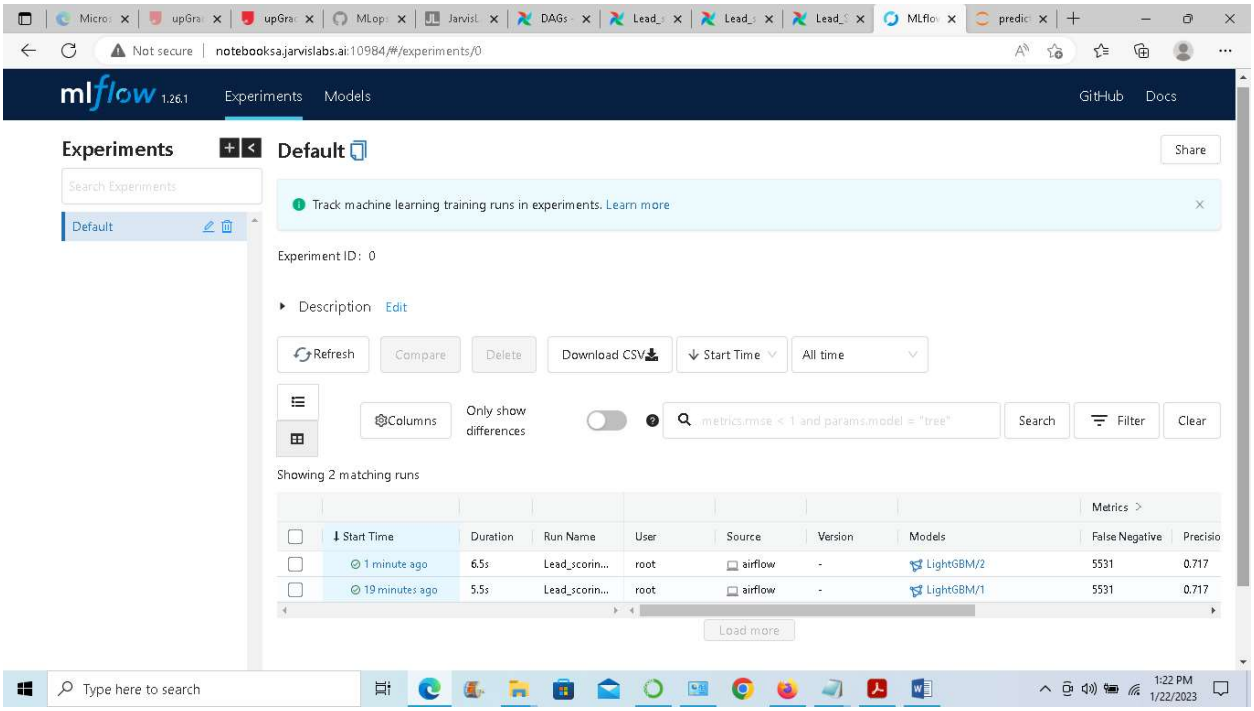

Not secure
notebooksa.jarvislabs.ai:10985/dags/Lead_scoring_inference_pipeline/graph
Airflow
DAGs
Security
Browse
Admin
Docs
19:17 UTC
DAG: Lead_scoring_inference_pipeline
Inference pipeline of Lead Scoring system
success
Schedule: @hourly
Next Run: 2023-01-22, 19:00:00
Grid
Graph
Calendar
Task Duration
Task Tries
Landing Times
Gantt
Details
Code
Audit Log
2023-01-22T19:13:09Z
Runs
25
Run
manual__2023-01-22T19:13:08.235115+00:00
Layout
Left > Right
Update
Find Task...
PythonOperator
deferred
failed
queued
running
scheduled
skipped
success
up_for_reschedule
up_for_retry
upstream_failed
no_status
Auto-refresh
encoding_categorical_variables
checking_input_features
generating_models_prediction
checking_model_prediction_ratio
Type here to search
1:17 PM
1/22/2023

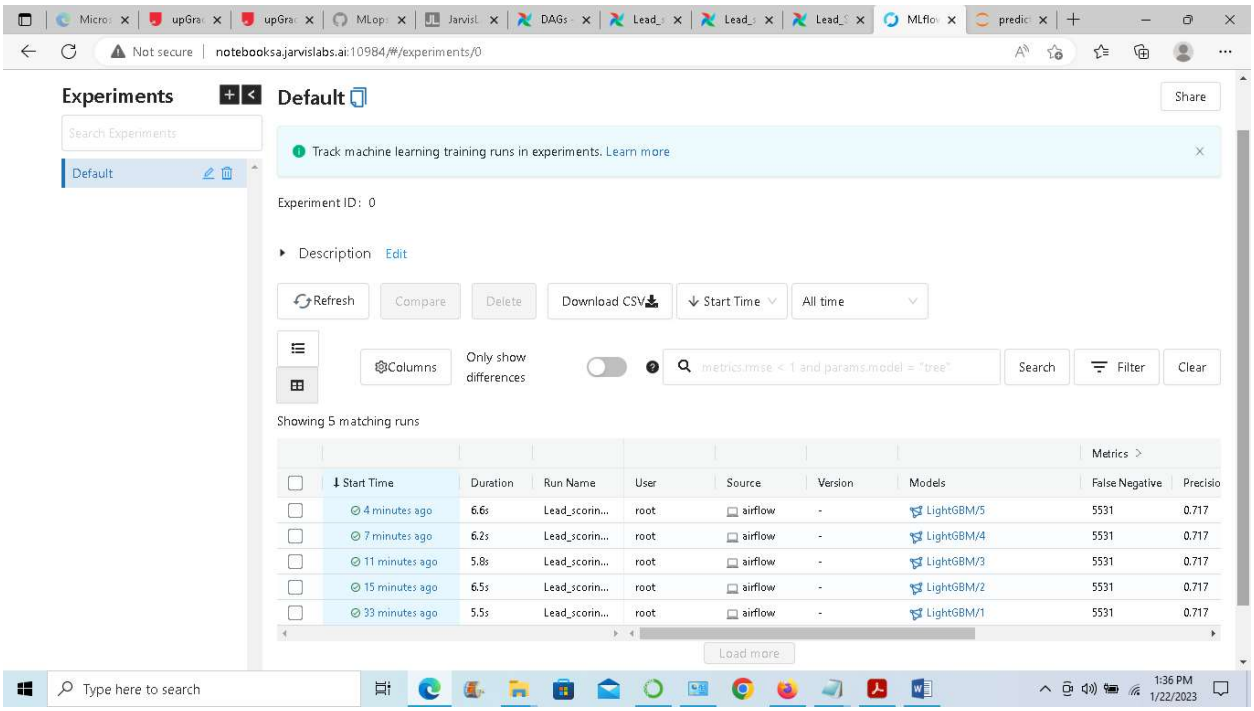
MLFlows UI:

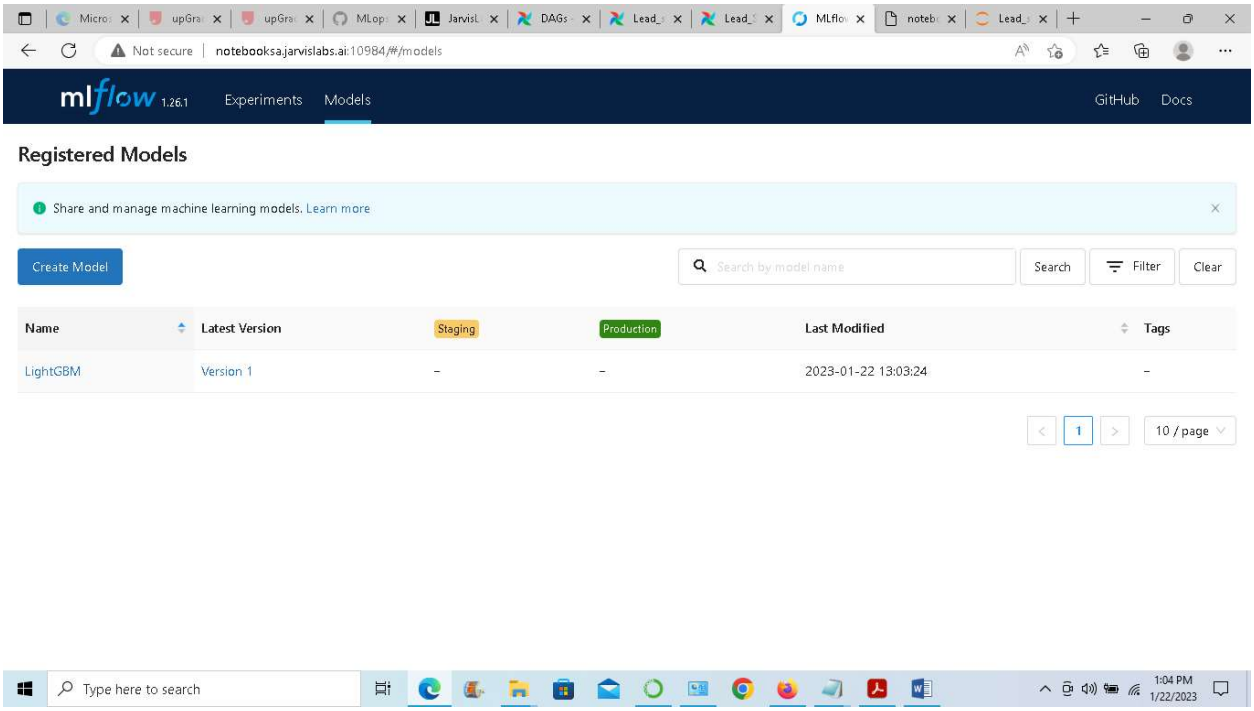

https://notebooksa.jarvislabs.ai/066/lab/tree/airflow/dags/Lead_scoring_training_pipeline
File Edit View Run Kernel Git Tabs Settings Help
Terminal 1
Terminal 3
utils.py
lead_scoring_training_p
Terminal 2
Filter files by name
/ dags / Lead_scoring_training_pipeline /
Name
Last Modified
mlruns 11 minutes ago
constants.py 20 hours ago
lead_scoring_training_pipeline.py 20 hours ago
TestTrainingPipeline.ipynb 15 minutes ago
utils.py 3 minutes ago
root@831c5f69e120:~# mlflow server --backend-store-uri='sqlite:////home/airflow/dags/Lead_scoring_data_pipeline/Lead_scoring_mlflow_production.db' --default-artifact-root="/airflow/dags/Lead_scoring_training_pipeline/mlruns/" --port=6006 --host=0.0.0.0
2023/01/22 18:51:31 INFO mlflow.store.db.utils: Creating initial MLflow database tables...
2023/01/22 18:51:31 INFO mlflow.store.db.utils: Updating database tables
INFO [alembic.runtime.migration] Context impl SQLiteImpl.
INFO [alembic.runtime.migration] Will assume non-transactional DDL.
INFO [alembic.runtime.migration] Running upgrade -> 451aebb31d03, add metric step
INFO [alembic.runtime.migration] Running upgrade 451aebb31d03 -> 90e64c465722, migrate user column to tags
INFO [alembic.runtime.migration] Running upgrade 90e64c465722 -> 181f10493468, allow nulls for metric values
INFO [alembic.runtime.migration] Running upgrade 181f10493468 -> df50e92ffc5e, Add Experiment Tags Table
INFO [alembic.runtime.migration] Running upgrade df50e92ffc5e -> 7ac759974ad8, Update run tags with larger limit
INFO [alembic.runtime.migration] Running upgrade 7ac759974ad8 -> 89d4b8295536, create latest metrics table
INFO [89d4b8295536_create_latest_metrics_table_py] Migration complete!
INFO [alembic.runtime.migration] Running upgrade 89d4b8295536 -> 2b4d017a5e9b, add model registry tables to db
INFO [2b4d017a5e9b_add_model_registry_tables_to_db_py] Adding registered_models and model_versions tables to database.
INFO [2b4d017a5e9b_add_model_registry_tables_to_db_py] Migration complete!
INFO [alembic.runtime.migration] Running upgrade 2b4d017a5e9b -> cfd24bdc0731, Update run status constraint with killed
INFO [alembic.runtime.migration] Running upgrade cfd24bdc0731 -> 0a8213491aaa, drop_duplicate_killed_constraint
INFO [alembic.runtime.migration] Running upgrade 0a8213491aaa -> 728d730b5ebd, add registered model tags table
INFO [alembic.runtime.migration] Running upgrade 728d730b5ebd -> 27a6a02d2cf1, add model version tags table
INFO [alembic.runtime.migration] Running upgrade 27a6a02d2cf1 -> 84291f40a231, add run_link to model_version
INFO [alembic.runtime.migration] Running upgrade 84291f40a231 -> a8c4a736bde6, allow nulls for run_id
INFO [alembic.runtime.migration] Running upgrade a8c4a736bde6 -> 39d1c3be5f05, add_is_nan_constraint_for_metrics_tables_if_necessary
INFO [alembic.runtime.migration] Running upgrade 39d1c3be5f05 -> c48cb773bb87, reset_default_value_for_is_nan_in_metrics_table_for_mysql
INFO [alembic.runtime.migration] Running upgrade c48cb773bb87 -> bd07f7e963c5, create index on run_uuid
INFO [alembic.runtime.migration] Context impl SQLiteImpl.
INFO [alembic.runtime.migration] Will assume non-transactional DDL.
[2023-01-22 18:51:31 +0000] [2913] [INFO] Starting gunicorn 20.1.0
[2023-01-22 18:51:31 +0000] [2913] [INFO] Listening at: http://0.0.0.0:6006 (2913)
[2023-01-22 18:51:31 +0000] [2913] [INFO] Using worker: sync
[2023-01-22 18:51:31 +0000] [2916] [INFO] Booting worker with pid: 2916
[2023-01-22 18:51:31 +0000] [2917] [INFO] Booting worker with pid: 2917
[2023-01-22 18:51:32 +0000] [2918] [INFO] Booting worker with pid: 2918
[2023-01-22 18:51:32 +0000] [2919] [INFO] Booting worker with pid: 2919
Simple
Terminal 3
Type here to search
1:00 PM
1/22/2023

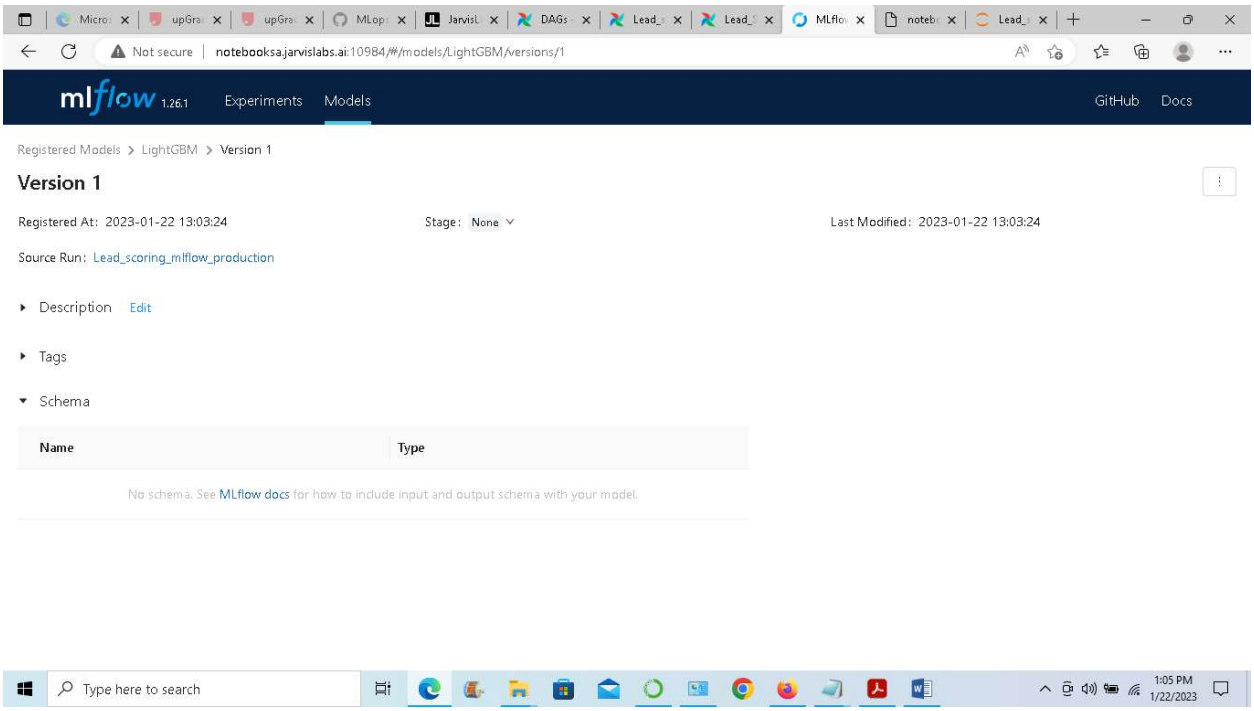

Not secure | notebooksa.jarvislabs.ai:10984/#/experiments/0
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Experiments
Search Experiments
Default
Default
Share
Track machine learning training runs in experiments. Learn more
Experiment ID: 0
Description Edit
Refresh
Compare
Delete
Download CSV
Start Time
All time
Columns
Only show differences
metrics.rmse < 1 and params.model = "tree"
Search
Filter
Clear
Showing 0 matching runs
Start Time
Duration
Run Name
User
Source
Version
Models
No runs yet. Learn more about how to create ML model training runs in this experiment.
Type here to search
12:52 PM
1/22/2023

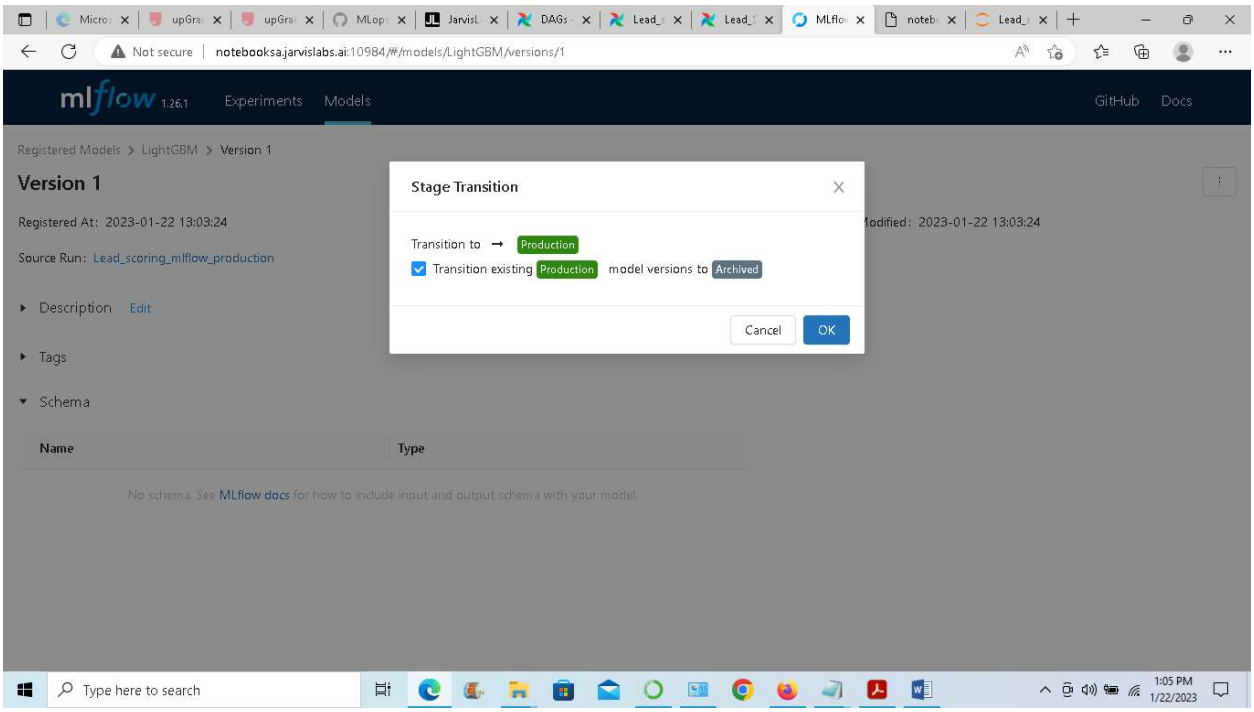

Not secure
notebooksa.jarvislabs.ai:10984/#/experiments/0
Experiments
Default
Share
Search Experiments
Default
Track machine learning training runs in experiments. Learn more
Experiment ID: 0
Description Edit
Refresh
Compare
Delete
Download CSV
Start Time
All time
Columns
Only show differences
metrics.rmse < 1 and params.model = "tree"
Search
Filter
Clear
Showing 1 matching run
Metrics
Start Time
Duration
Run Name
User
Source
Version
Models
False Negative
Precisio
39 seconds ago
5.5s
Lead_scorin...
root
airflow
-
LightGBM/1
5531
0.717
Load more
Type here to search
1:04 PM
1/22/2023

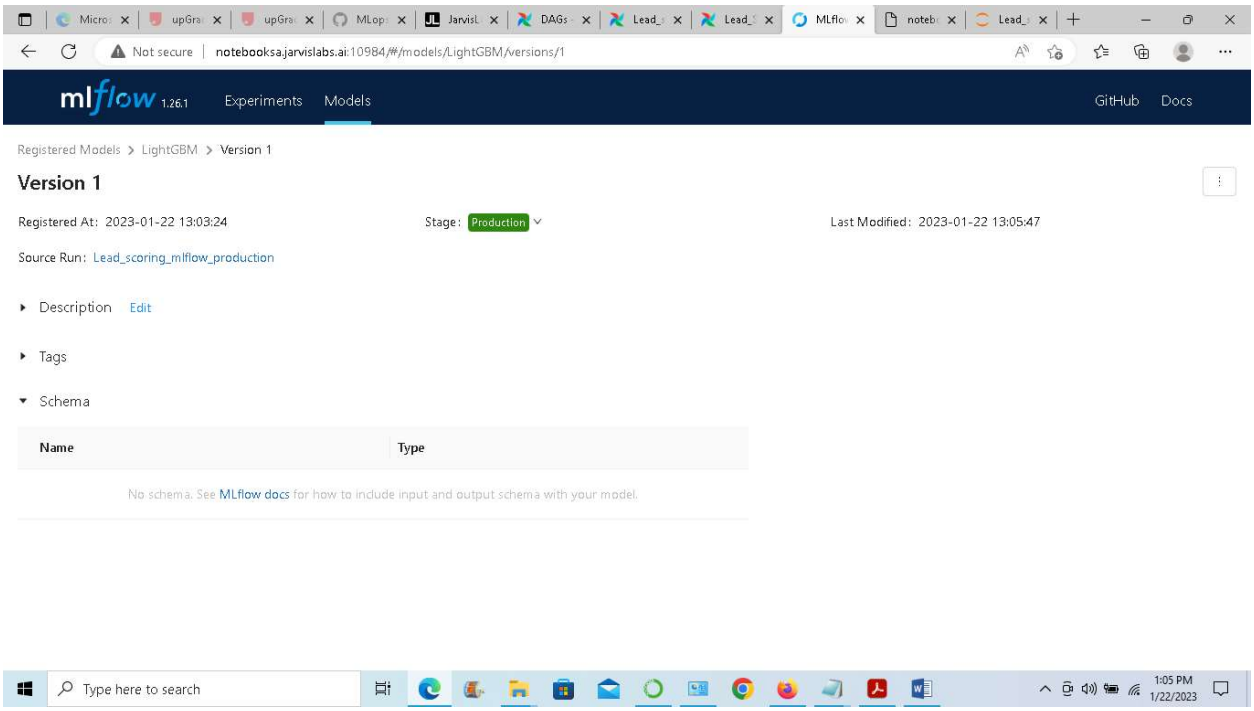

Not secure
notebooksa.jarvislabs.ai:10984/#/experiments/0
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Experiments
Default
Share
Search Experiments
Default
Track machine learning training runs in experiments. Learn more
Experiment ID: 0
Description Edit
Refresh
Compare
Delete
Download CSV
Start Time
All time
Columns
Only show differences
metrics.rmse < 1 and params.model = "tree"
Search
Filter
Clear
Showing 2 matching runs
Metrics
Start Time
Duration
Run Name
User
Source
Version
Models
False Negative
Precisio
1 minute ago
6.5s
Lead_scorin...
root
airflow
-
LightGBM/2
5531
0.717
19 minutes ago
5.5s
Lead_scorin...
root
airflow
-
LightGBM/1
5531
0.717
Load more
Type here to search
1:22 PM
1/22/2023

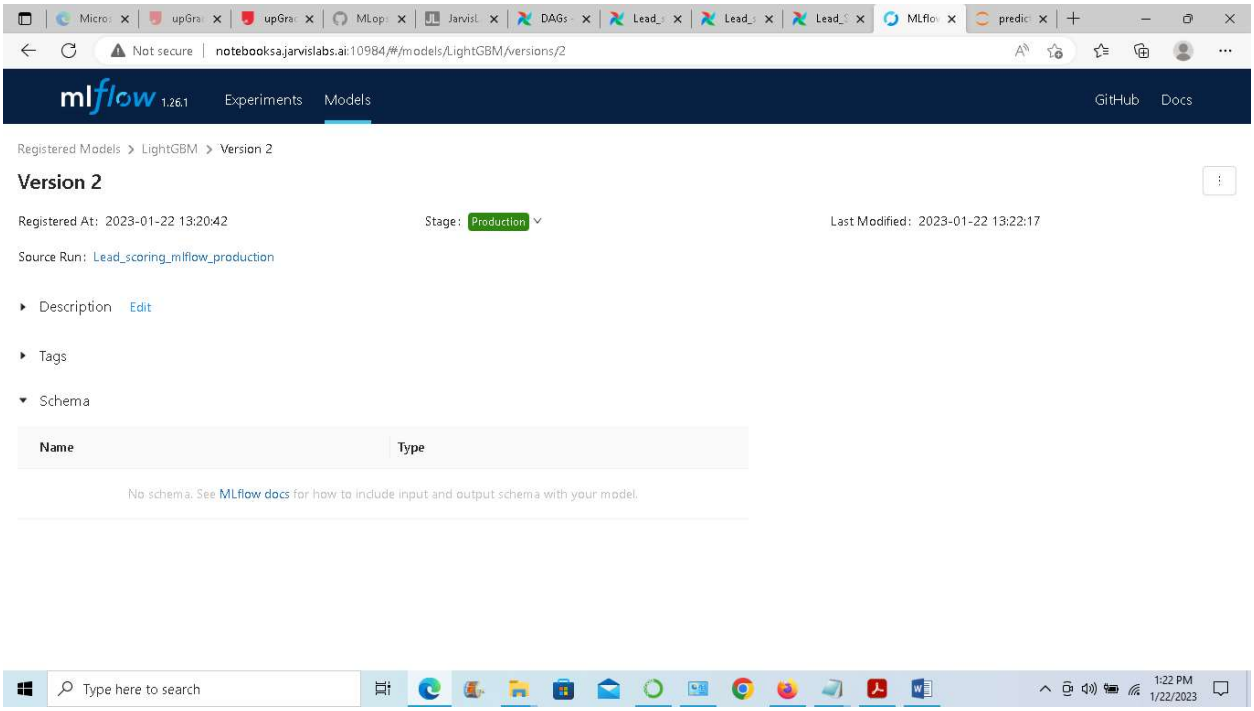

Not secure
notebooksa.jarvislabs.ai:10984/#/experiments/0
Experiments
Default
Share
Search Experiments
Default
Track machine learning training runs in experiments. Learn more
Experiment ID: 0
Description Edit
Refresh
Compare
Delete
Download CSV
Start Time
All time
Columns
Only show differences
metrics.rmse < 1 and params.model = "tree"
Search
Filter
Clear
Showing 5 matching runs
Metrics
Start Time
Duration
Run Name
User
Source
Version
Models
False Negative
Precisio
4 minutes ago
6.6s
Lead_scorin...
root
airflow
-
LightGBM/5
5531
0.717
7 minutes ago
6.2s
Lead_scorin...
root
airflow
-
LightGBM/4
5531
0.717
11 minutes ago
5.8s
Lead_scorin...
root
airflow
-
LightGBM/3
5531
0.717
15 minutes ago
6.5s
Lead_scorin...
root
airflow
-
LightGBM/2
5531
0.717
33 minutes ago
5.5s
Lead_scorin...
root
airflow
-
LightGBM/1
5531
0.717
Load more
Type here to search
1:36 PM
1/22/2023

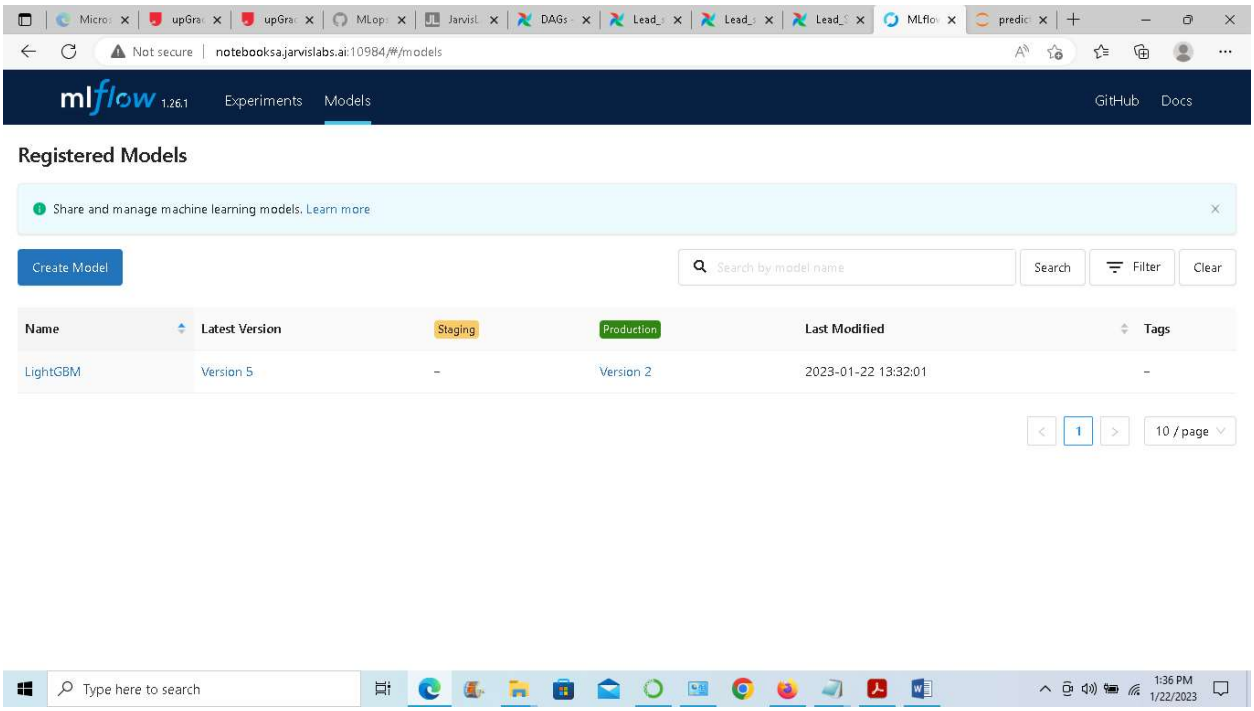

Not secure
notebooksa.jarvislabs.ai:10984/#/models
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Registered Models
Share and manage machine learning models. Learn more
Create Model
Search by model name
Search
Filter
Clear
Name
Latest Version
Staging
Production
Last Modified
Tags
LightGBM
Version 1
-
-
2023-01-22 13:03:24
-
1
10 / page
Type here to search
1:04 PM
1/22/2023


Not secure
notebooksa.jarvislabs.ai:10984/#/models/LightGBM/versions/1
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Registered Models > LightGBM > Version 1
Version 1
Registered At: 2023-01-22 13:03:24
Stage: None
Last Modified: 2023-01-22 13:03:24
Source Run: Lead_scoring_mlflow_production
Description
Edit
Tags
Schema
Name
Type
No schema. See MLflow docs for how to include input and output schema with your model.
Type here to search
1:05 PM
1/22/2023

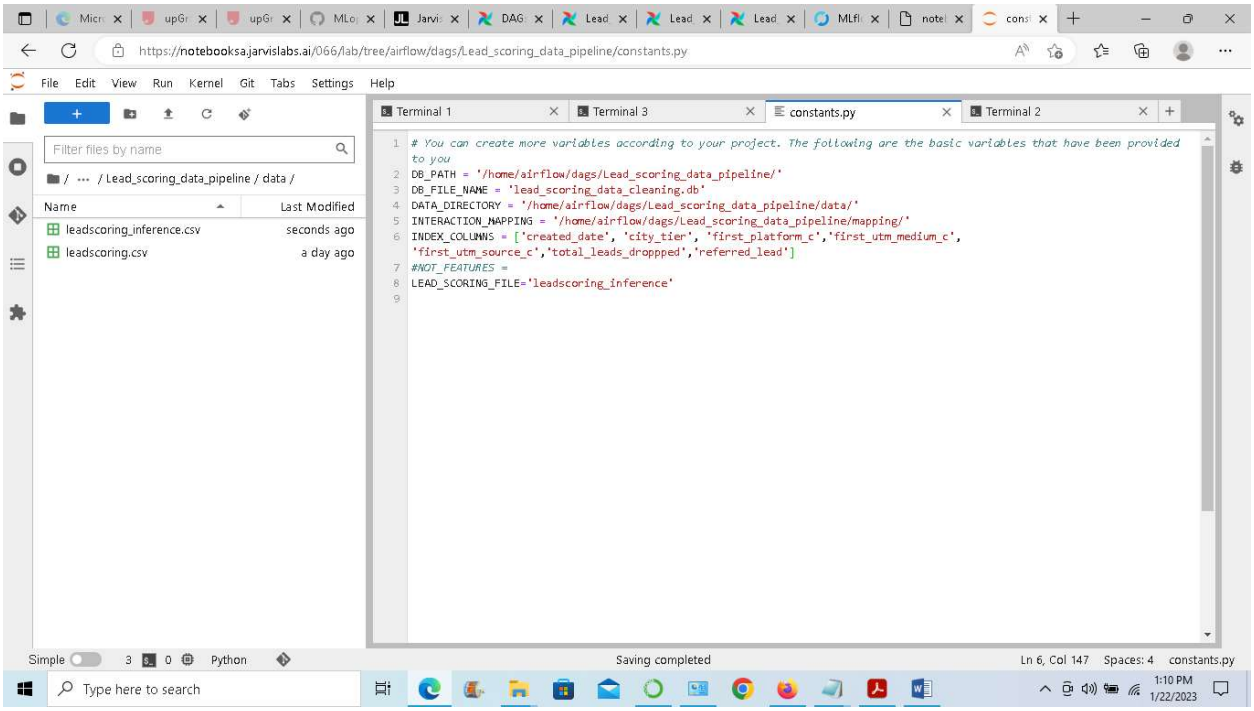

Not secure
notebooksa.jarvislabs.ai:10984/#/models/LightGBM/versions/1
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Registered Models > LightGBM > Version 1
Version 1
Registered At: 2023-01-22 13:03:24
Source Run: Lead_scoring_mlflow_production
Description
Edit
Tags
Schema
Name
Type
No schema. See MLflow docs for how to include input and output schema with your model.
Stage Transition
Transition to → Production
Transition existing Production model versions to Archived
Cancel
OK
Type here to search
1:05 PM
1/22/2023


Not secure
notebooksa.jarvislabs.ai:10984/#/models/LightGBM/versions/1
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Registered Models > LightGBM > Version 1
Version 1
Registered At: 2023-01-22 13:03:24
Stage: Production
Last Modified: 2023-01-22 13:05:47
Source Run: Lead_scoring_mlflow_production
Description Edit
Tags
Schema
Name
Type
No schema. See MLflow docs for how to include input and output schema with your model.
Type here to search
1:05 PM
1/22/2023

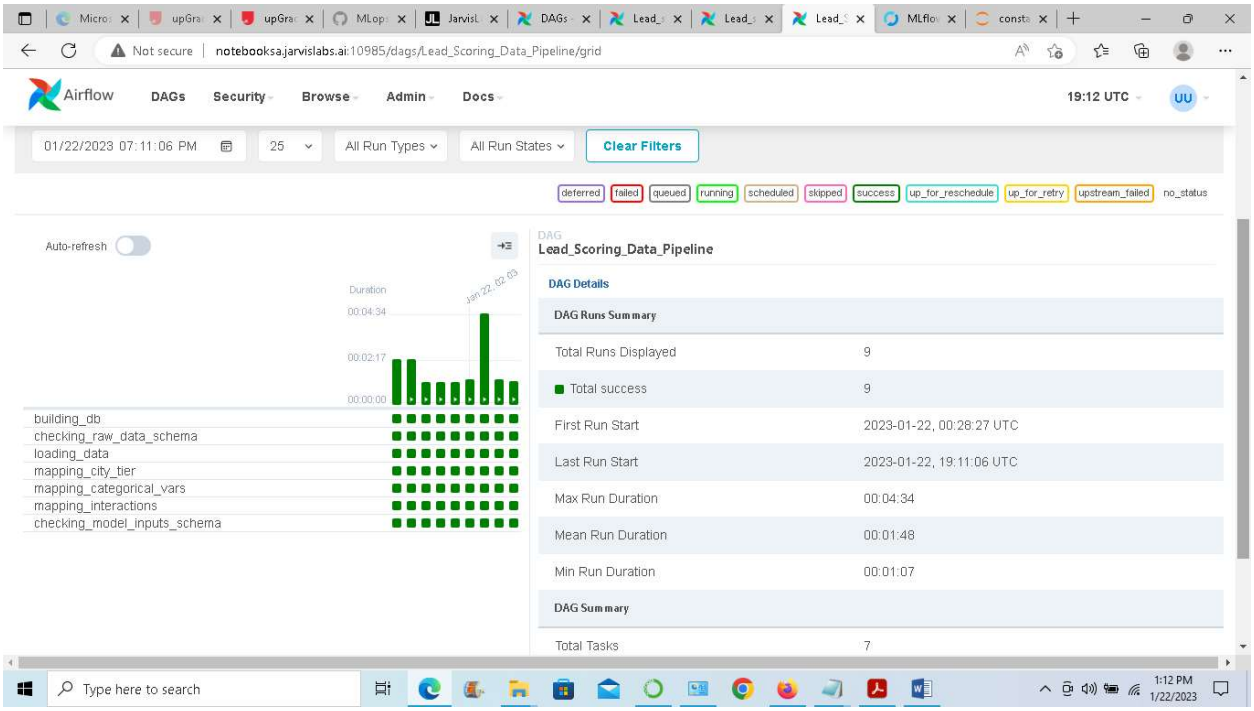

Not secure
notebooksa.jarvislabs.ai:10984/#/models/LightGBM/versions/2
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Registered Models > LightGBM > Version 2
Version 2
Registered At: 2023-01-22 13:20:42
Stage: Production
Last Modified: 2023-01-22 13:22:17
Source Run: Lead_scoring_mlflow_production
Description Edit
Tags
Schema
Name
Type
No schema. See MLflow docs for how to include input and output schema with your model.
Type here to search
1:22 PM
1/22/2023



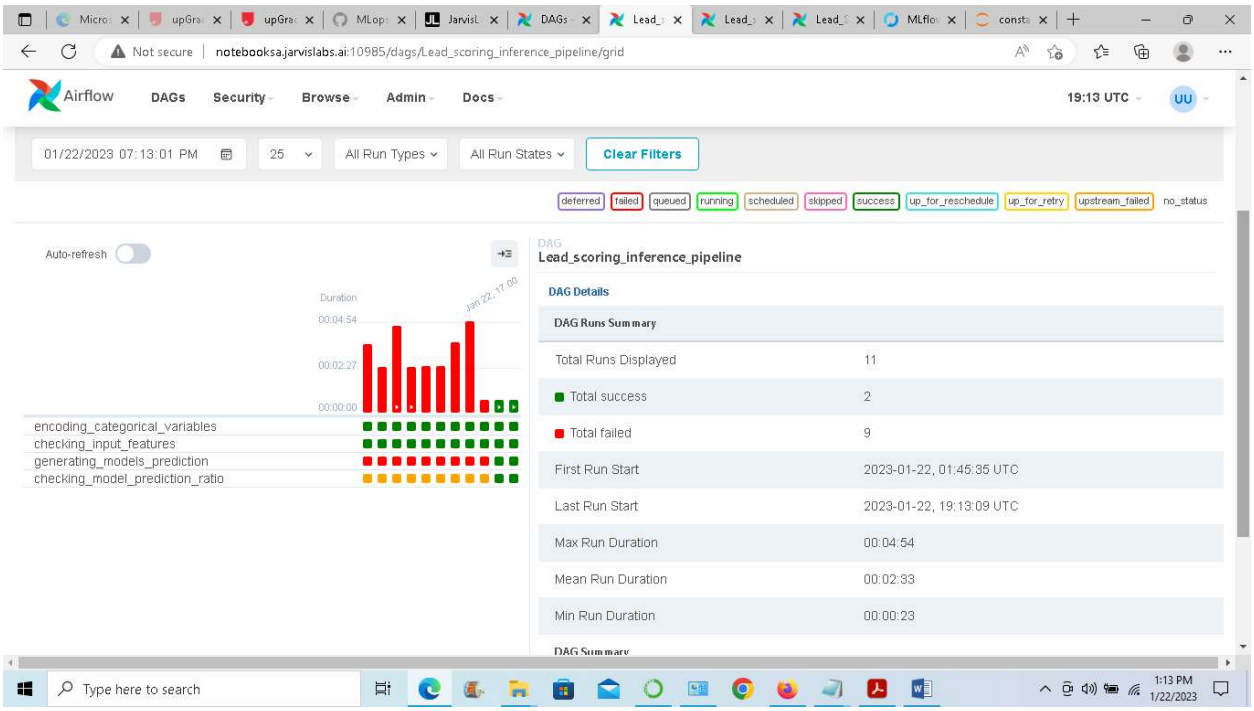
After pointing to Inference file

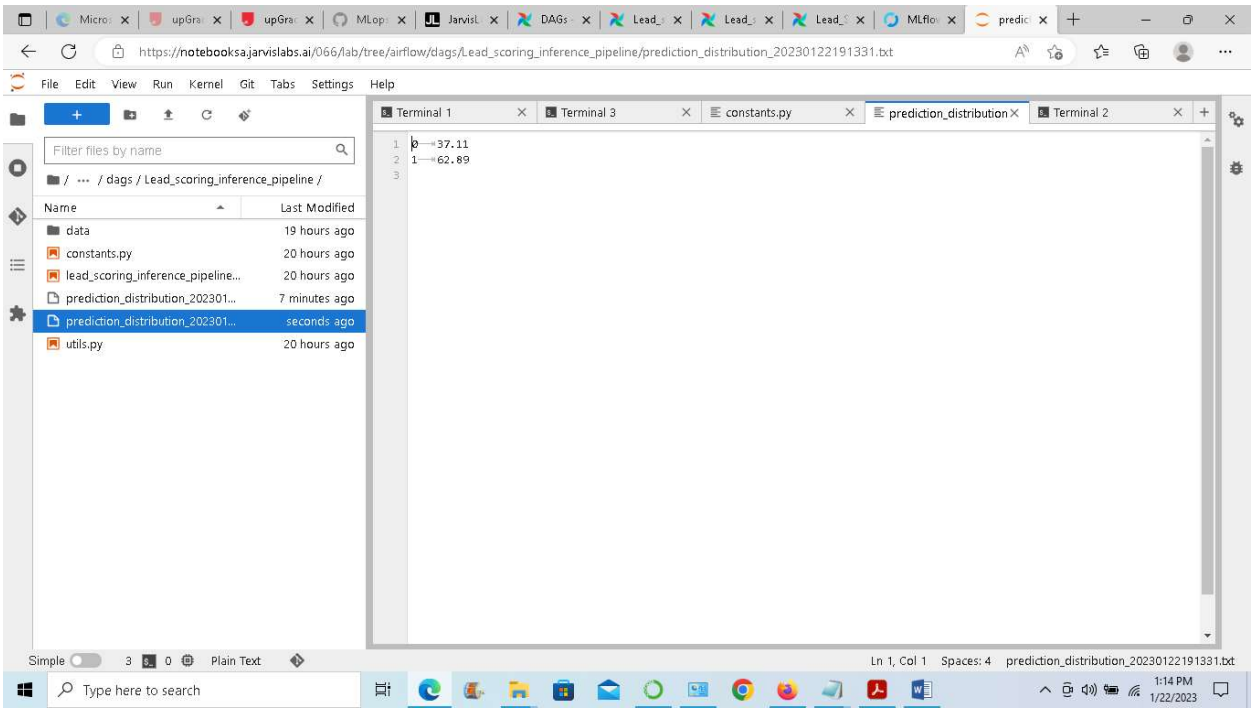


Data Pipeline



Inference Pipeline


Not secure notebooksa.jarvislabs.ai:10985/dags/Lead_scoring_inference_pipeline/grid
Airflow
DAGs
Security
Browse
Admin
Docs
19:13 UTC
01/22/2023 07:13:01 PM
25
All Run Types
All Run States
Clear Filters
deferred
failed
queued
running
scheduled
skipped
success
up_for_reschedule
up_for_retry
upstream_failed
no_status
Auto-refresh
Duration
00:04:54
00:02:27
00:00:00
encoding_categorical_variables
checking_input_features
generating_models_prediction
checking_model_prediction_ratio
DAG
Lead_scoring_inference_pipeline
DAG Details
DAG Runs Summary
Total Runs Displayed 11
Total success 2
Total failed 9
First Run Start 2023-01-22, 01:45:35 UTC
Last Run Start 2023-01-22, 19:13:09 UTC
Max Run Duration 00:04:54
Mean Run Duration 00:02:33
Min Run Duration 00:00:23
DAG Summary
Type here to search
1:13 PM
1/22/2023


https://notebooksa.jarvislabs.ai/066/lab/tree/airflow/dags/Lead_scoring_inference_pipeline/prediction_distribution_20230122191331.txt
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ dags / Lead_scoring_inference_pipeline /
Name
Last Modified
data 19 hours ago
constants.py 20 hours ago
lead_scoring_inference_pipeline... 20 hours ago
prediction_distribution_202301... 7 minutes ago
prediction_distribution_202301... seconds ago
utils.py 20 hours ago
Terminal 1
Terminal 3
constants.py
prediction_distribution
Terminal 2
0 37.11
1 62.89
Simple
Plain Text
Ln 1, Col 1 Spaces: 4 prediction_distribution_20230122191331.txt
Type here to search
1:14 PM
1/22/2023